8-2010

Direct visualization of cold denaturation of cytochrome-C Using Small Angle X-Ray Scattering

Margaret Elmer
DePaul University, margaret.m.elmer@gmail.com

DIRECT VISUALIZATION OF COLD DENATURATION OF CYTOCHROME-C USING SMALL ANGLE X-RAY SCATTERING

A Thesis

Presented in

Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

August 1, 2010

BY

Margaret Elmer

PHYSICS DEPARTMENT

College of Liberal Arts and Sciences

DePaul University

Chicago, Illinois

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

We present the first comprehensive study of the cold denaturation of proteins using Small Angle X-Ray Scattering. The radius of gyration of equine Cytochrome-C is determined for varying salt and denaturant conditions at temperatures ranging from -25 to 60 degrees Celsius. Radius of gyration measurements are then compared to a theoretical protein folding theory using by using a two-state population model to relate thermodynamic quantities to physical measurements. The incorporation of a temperature-dependent pH and solvent dielectric constant is critical to incorporating the electrostatic interactions of both the protein and the solution over this broad temperature range and properly predict the observed protein stability from sequence. At suitable conditions, the protein can be made to increase in size by nearly 9 Angstroms (over 60% of its native radius of gyration) when dropped in temperature from 0 to -25 degrees Celsius. This result is promising for future studies of ultrafast protein folding using time-resolved SAXS where initially cold denatured protein will be suddenly jumped in temperature by an infrared laser to initiate folding. Further, this work validates modifications made to existing protein folding theory.

ACKNOWLEDGEMENTS

To all those who put more effort than I deserved into this work, I offer my gratitude.

For my Mother, Father and Steven Langel

A special thanks to all contributors as Sector 18 APS

Dr Landahl

Dr Sarma

Dr Pando

The Rice Lab

The Lin Lab

The Grueble Lab

Joseph Marcus

Christopher Asta

And every other laboratory collaborator in the Landahl Lab.

CHAPTER 1

Introduction

Section 1.1: The Protein Folding Problem

All organisms need proteins to live. Proteins serve many functions in the body. Constructed by ribosomes inside the cell, proteins are formed as specific sequences of amino acids indirectly encoded by DNA (Nelson 2005). The protein-folding problem poses the question of how strands or sequences of amino acids fold into multi-dimensional, functional, stable structures in the body. This is the protein-folding problem. This thesis addresses a fundamental aspect of the problem: specifically, what physical properties are required to stabilize the folded state of proteins? Here we demonstrate experimentally that manipulation of several different parameters may lead to instability and unfolding of a protein. Explanation of this data requires a model that accounts for both the physical characteristics of the protein itself as well as its environment.

Section 1.2: Cellular Folding

As proteins are constructed in the ribosome, folding begins to occur. The amino acids manipulate each other on both local and global scales. Electrostatic forces pull amino acids close in sequence into approximate location. A final conformation is reached. The conformation that is stable and functional is known as the native conformation or state. Proteins do not always assume their native state after production by a ribosome. In this case, the final conformation does not meet the requirements of stability and or functionality. Further, the lowest free energy state is not necessarily the native state of the protein.

Proteins in the cell naturally unfold and fold spontaneously. The environment of the protein affects this unfolding and folding around an equilibrium population distribution. A shift of the protein population to a greater proportion in the unfolded state at equilibrium is known as denaturation. In the following experiments, the proportion of folded vs. unfolded protein is determined by measuring the average size of a distribution of proteins at a particular set of conditions. The observation of protein denaturation through variation of physical parameters highlights the processes that drive protein folding.

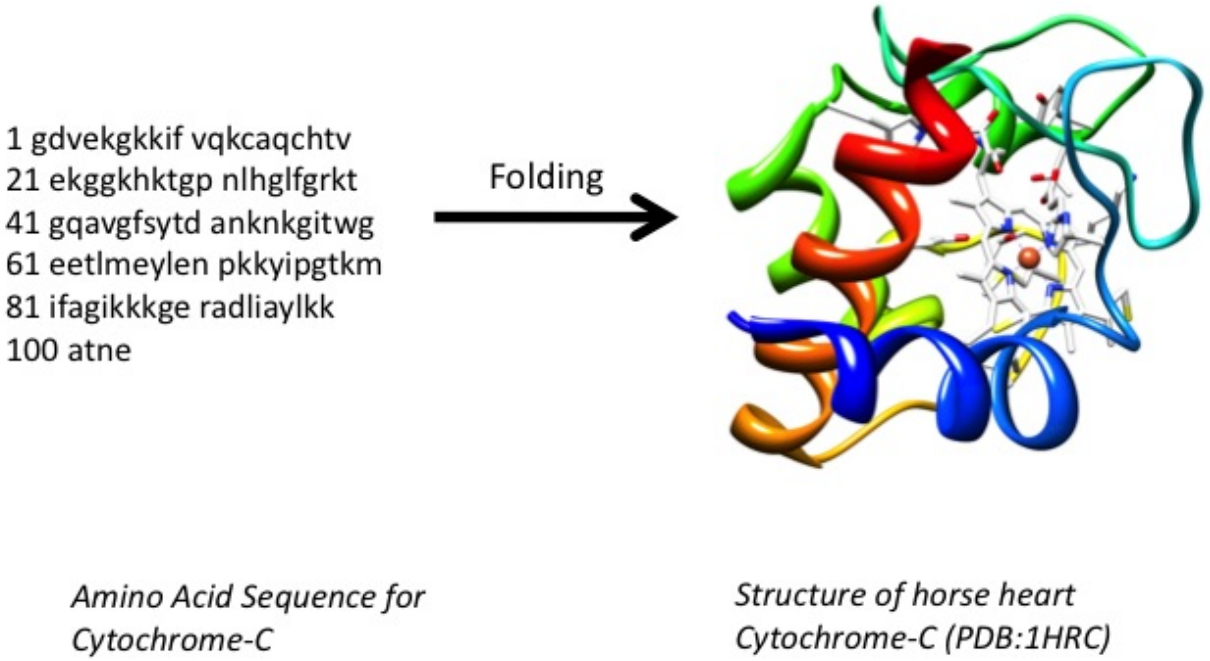


Figure 1.1: The protein folding problem for the Cytochrome-C sequence as it is composed by the ribosome and then finds its way to its most stable conformation.

Section 1.3: Previous Studies

Extensive work has been done to aid in the understanding of how proteins develop into final functional, stable structures from their ribosome assembled amino acid sequences. Techniques have been developed to use visible light and near visible light to study proteins as they fold (Press 2000). The use of Nuclear Magnetic Resonance or NMR has also been used to study protein structure of folded and unfolded proteins. NMR, when done on various fragments of the same protein, can give insight into secondary structure of a protein. NMR is time consuming and works on a time scale much longer than the average time it takes for a protein to fold. Fluorescence testing has also been implemented to better understand how proteins fold. Known as Förster Resonance Energy Transfer or FRET, is a fluorescence exposure testing that behaves as a local protein probe by providing information about fluorescence absorption for distinct regions of the protein. If it is known where the fluorescent responsive portions of the protein occur in the amino acid sequence, FRET may aid in measurement of size and characterization of secondary structures of the protein in study (Silvius 2006). Because FRET is capable of providing information about the size of a protein, it can also be used as global probe. A third standard technique of studying protein folding is Circular Dichroism or CD. To study protein secondary structure, circular dichroism works by exposing a protein sample to UV light between 200 and 250 nanometers. Alpha helices and beta sheets in proteins respond to circularly polarized light with very different absorption patterns in this wavelength range (Rodger 1997). Exposing a protein to CD can verify alpha helices or beta sheets predicted by associated genetic sequences. In the case of a randomly folded protein that has neither alpha helices nor beta sheets,

as is generally the case of a denatured protein, CD is able to function as a local probe in determining secondary structure loss or change. These studies have the capability of investigating portions of the protein and render local structural information. In some cases, these studies may also be able to provide information about the global structure of a protein. Research suggests that protein folding is not only dependent on local bonding but long range bonding as well. Global or tertiary structure changes in the protein need to be studied to better understand how proteins find the most stable and functional or native state.

Theoretical and computational investigations have also been conducted in an attempt to explain how a protein starts as an amino acid chain and find its way into lowest functional free energy conformation and or native state. One such study published in 1998 theorized that a protein chain of 100 amino acids would have 10^{18} potential conformations that it could assume (Dobson 1998). Using the approximation that a conformational change takes 10^{-11} seconds, a computer program could be run to calculate every possible conformation. Unfortunately, testing every conformation possible using this change in conformation time scale would require a computation time of 10^{30} years (Dobson 1998). Twelve years later, it is now possible to go on the Internet and find a program that will calculate the lowest free energy state of a given protein on a personal computer in just two days! But the average protein takes less than a second to find this conformation. Through chemical manipulation, fragments of proteins have been studied to better understand how each segment finds its most thermodynamically stable, active conformation. Such studies have revealed that certain amino acid sequences inherently lead to stable three-dimensional structures. Computational studies like those found on Pubmed or through the Protein Data Bank, also known as the PDB, have utilized these findings for faster calculations of protein structure (Press 2000). Generally, these programs utilize a protein folding algorithm that starts with the assumption that once out of the ribosome, the amino acid sequence hydrophobically collapses. In this process, the amino acid sequence collapses a random structure with the hydrophobic portions protected from the host environment. Hydrophobic collapse removes a large portion of the possible conformations and reduces the number of calculations that need to be performed. From genomic studies, several secondary structures are known to be directly associated with specific amino acid sequences. Computation efforts like those being discussed, then incorporate these secondary structure relations to further eliminate possible conformations. Finally, tertiary structures are considered for a given protein's amino acid sequence to again eliminate possible conformations (Pain 1994). Once these three characterizations are considered, protein structure can be computed. This algorithm has not been experimental validated for all proteins. In fact, other algorithms exist that also attempt to predict protein conformation. Such algorithms are more time consuming and thus more difficult to use. Studying the global change in structure may be helpful in validating such an algorithm and understanding a protein's ability to find its native state in such a short amount of time.

Since the early 1940's, Small Angle X-ray Scattering, or SAXS, has been developed as a technique for determining the size and approximate shape of small objects (Glatter 1982). Given a sufficiently bright x-ray source, particles of less than 2 microns can be exposed to an x-ray beam and their scattering pattern can be processed and analyzed to determine their approximate size and shape. SAXS can be used to study a protein's global structure in liquid. Because the scattering analysis method only measures electron density, proteins that change size as they unfold are good candidates for this technique. As the protein unfolds, its electron density should decrease. Small angle x-ray scattering, directly determines the size of the protein, thus answering

the question of whether it is in its naturally occurring or native conformation or its unfolded or denatured conformation.

Section 1.4: Protein Unfolding

At physiological conditions most proteins are folded. Varying the physical parameters of the protein's environment can weaken short and long-range cohesive forces of the protein and unfold it. This artificially induced unfolding of a protein is known as denaturation. Several known chemicals can be added to a protein's environment to induce protein denaturation. In this experiment, Guanidine HCl, GuHCl, is used at varying concentrations to denature the protein population (Kumar 2006). Varying the pH of a solution also acts to denature a protein by weakening the local and global forces holding the native structure together (Kumar 2006). Thermodynamically favorable conditions may also act to denature a protein (Privalov 1974), which will be the subject of Chapter 2. These denaturing factors are used in conjunction to unfold Cytochrome-C in this experiment.

Section 1.5: Slowly Unfolding the Protein to answer the Protein folding Question

The physical parameters of a protein's environment may work to render a protein unstable and lead to warm and cold protein denaturation. For a well studied protein like Cytochrome-C, it is known that drastic structural change occurs as the protein denatures. In the native state, Cytochrome-C can be approximated as a sphere. As Cytochrome-C unfolds, it has been predicted to pass through two known intermediates elongating into its final denatured state. At each stage of unfolding, the radius of gyration associated with the stage increase. Using SAXS, measurement of the radius of gyration of Cytochrome-C with high accuracy is possible. These measurements permit the answering of the question of what physical parameters control protein conformation.

Chapter 2

Cold Denaturation

Thermodynamics can be used to explain the behavioral changes of a protein in solution. A formal development is needed to see how the physical parameters of a protein solution change with varying temperature. The simplest determination of protein unfolding is a change in secondary and tertiary structure. Structural change is indicative of state change. From thermodynamics, state change can be directly described through a study of the Gibbs and the Helmholtz free energies of the system. To study the thermodynamic energies of the system, specific heat with respect to the pressure, the entropy and enthalpy should be considered when studying such a system. Starting from first principles, the relationships between each of these individual elements can be derived. Building a formal understanding of how the system changes with temperature will cement into place, the validity of a thermodynamic approach to macromolecular systems like proteins.

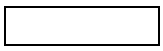
This chapter will discuss the basic thermodynamic prediction of cold denaturation. A full theoretical derivation will be implemented starting with the consideration of the thermodynamics of the system. Using electrodynamics, a complete theoretical model for protein folding will be developed.

Section 2.1: Definition of Enthalpy

The definition of specific heat for constant pressure, p, is:

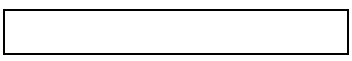
(2.1)

. Where H represents the enthalpy of the system, U represents the internal energy of the system and V, is the volume of the system. The enthalpy of a system is the measurement of heat transfer of a closed system under constant pressure. Equation (2.1) may also be written as:



(2.2)

. Where all values involving change can be calculated by subtracting an initial value from a final value:



(2.3)

or,

(2.4)

. Thus, the specific heat with respect to constant pressure has a direct correlation with enthalpy of the system. We define the temperature where the Enthalpy is zero as T_h, as illustrated in Fig. 2.1.

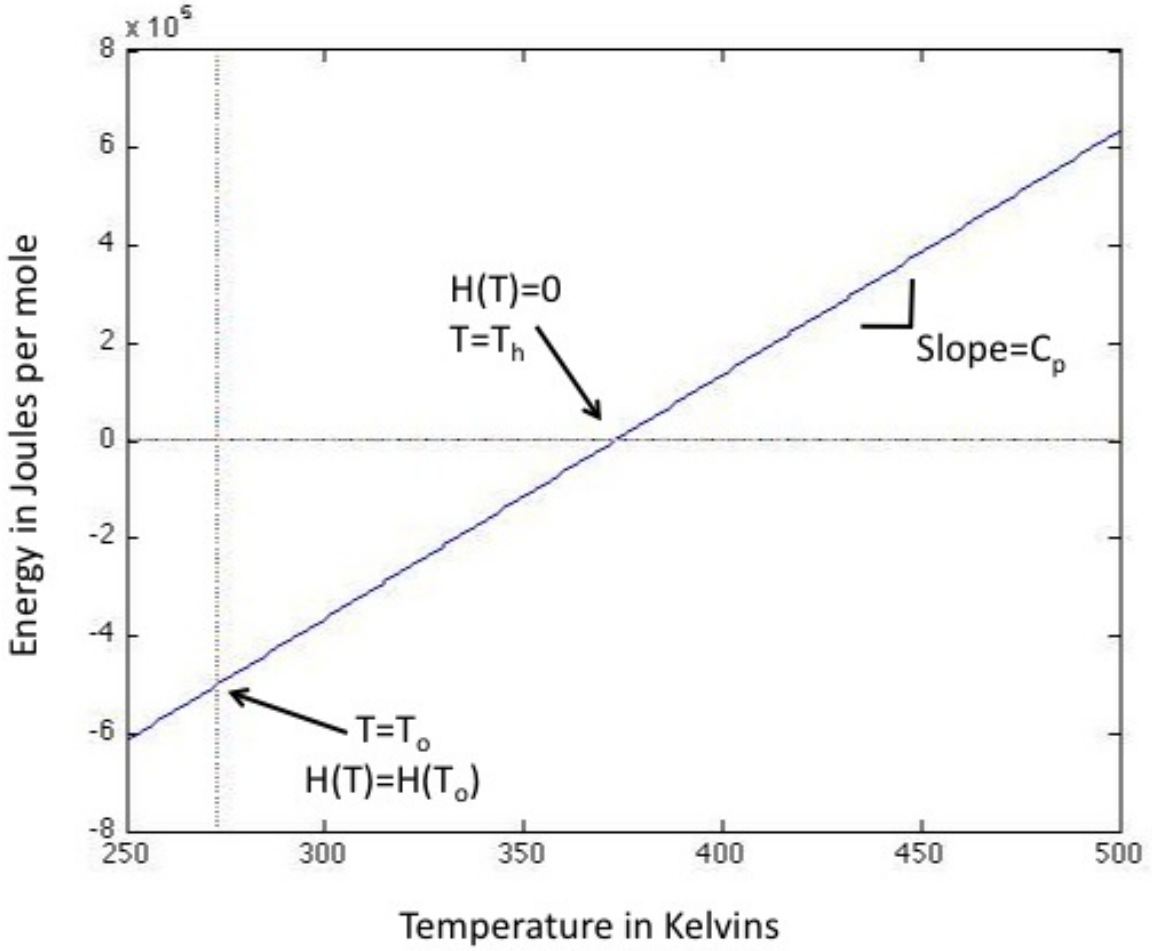


Figure 2.1: The heat of enthalpy in Joules per mole graphed verse temperature in Kelvins. Values of C_p, at 5000 Joules per mole per Kelvin, and T_h, at 373 Kelvins, are taken from the ideal thermal protein model for Cytochrome-C (Robertson 1997).

To the right of the x-intercept, when the temperature is greater than T_h, enthalpy is positive and will act as a barrier to overcome in system change. To the left of the x-intercept, when the temperature is less than T_h, enthalpy is negative, making system change more thermodynamically favorable. At T_h, the enthalpy has no effect on the system's ability to change. Temperature directly affects the enthalpy of a solution manipulates a system. Entropy also has the ability to manipulate a system. This will be shown next.

Section 2.2: Definition of Entropy

Entropy is the measurement of the microscopic level of order of a system and provides a formal definition of temperature. A well-ordered system has low entropy. As disorder increases, its value of entropy increases as well. Holding volume, V, and number density, N, constant, entropy, S, can be rewritten with respect to temperature and internal energy

(2.5)

or,

(2.6)

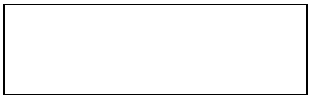
. Integrating both sides of the equation provides a relation that connects entropy to internal energy through temperature

(2.7)

. Using the value for specific heat of with respect to constant volume, the change in entropy may be rewritten as

(2.8)

or,



(2.9)

. With the assumption that $C_v \sim C_p$ for liquids, Eq. (2.9) may be approximated as:

(2.10)

. The temperature at which the entropy of the system vanishes will be denoted as T_s and is illustrated in Fig 2.2. This should not be incorrectly interpreted as the point at which the system is most ordered. Instead T_s is the temperature at which entropy is only dependent on the chemical structure of the system. At this temperature, the chain length of the protein comprising the system will be the only contributing factor for entropy.

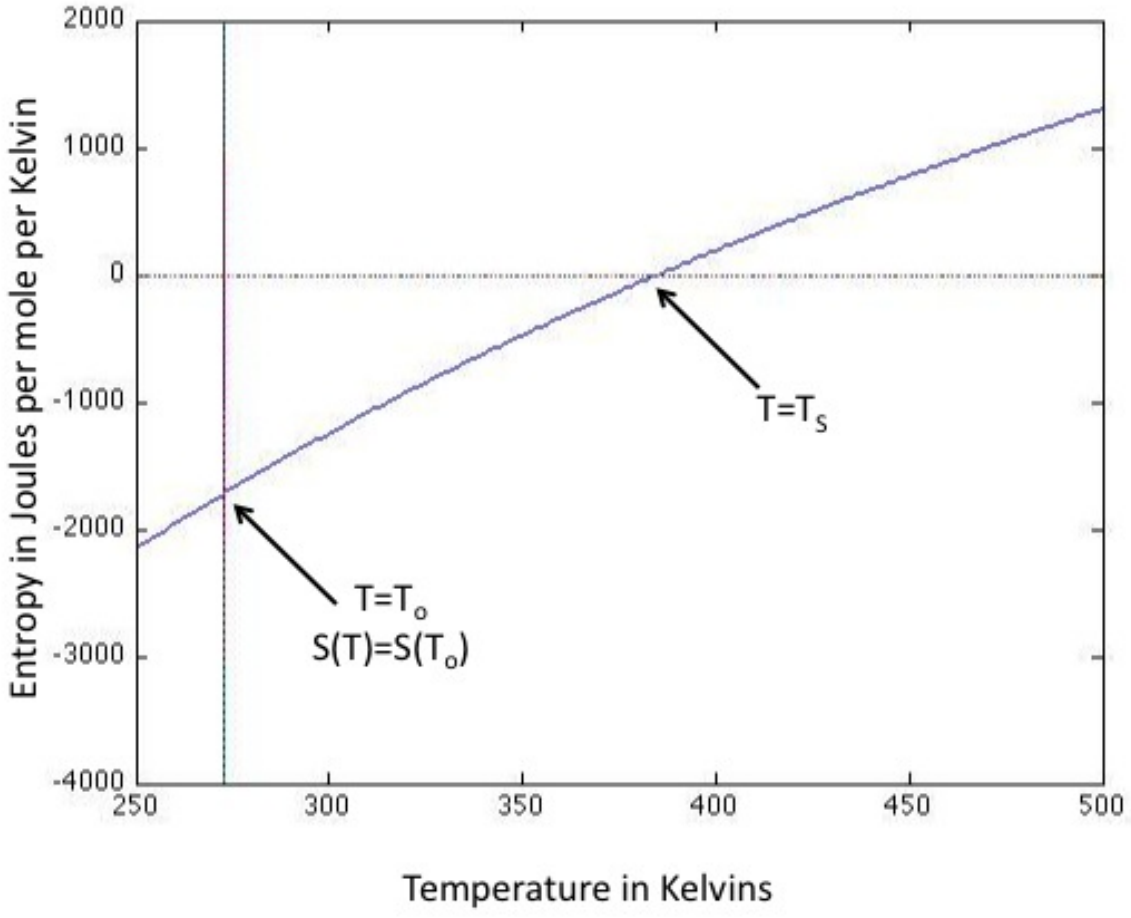


Figure 2.2: The entropy in Joules per mole per Kelvin graphed verse temperature in Kelvins. Values of C_p, at 5000 Joules per mole per Kelvin, and T_s, at 385 Kelvins, are taken from the ideal thermal protein model for Cytochrome-C (Robertson 1997).

Figure 2.2 demonstrates the relationship between entropy and temperature. At any location to the left of T_s, the host environment or system is unfavorable to change. Change is unlikely as in the case of the enthalpy. To the right of T_s, the system is favorable to change or willing to an increase in disorganization, satisfying the second law of thermodynamics. As temperature increases, both entropy and disorder increase. At T_s, when entropy is zero, the entropy has no affect on the system's ability to stabilize or destabilize a protein because it has no impact on the organization of the system.

Section 2.3: Relating Enthalpy and Entropy Through the Gibbs Equation

The Gibbs free energy, G, of a system is the amount of energy free to be used in a closed system without change in pressure. The Gibbs equation relates entropy and enthalpy thermodynamically to Gibbs free energy characterizing a system (Dill 2002):

(2.11)

When temperature is held constant, the change in the Gibbs free energy can be rewritten with respect to the change in enthalpy and the change and entropy.

(2.12)

The change in variables enthalpy and entropy must take into consideration the temperature at which the Gibbs free energy is calculated or measured and temperature dependence of each variable must now be reintroduced. The Gibbs free energy is not a function of temperature directly though. The Gibbs free energy is found by comparing independently prepared and measured solutions at different static temperatures. To represent these independent measurements, and remind the reader that this is theory is designed for static measurements and not for a thermodynamic process, the Gibbs free energy comparison at different temperatures for independently prepared solutions is denoted with a subscript temperature abbreviation. The Gibbs free energy at a given temperature found by comparing the enthalpy and entropy as functions of temperature,

(2.13)

. Substituting Eq. (2.4) and Eq. (2.10) into the expression for the change in Gibbs free energy, $(\Delta G)_T$, yields:

(2.14)

. Here, the temperature T_o, represents the initial or reference temperature measurement for entropy or enthalpy. The Gibbs equation can be simplified through the choice of the arbitrary reference temperature T_o.

First, choose T_0 to be the temperature at such that:

(2.15)

. Now the change in entropy may be written simply as:

(2.16)

. Likewise, the change in enthalpy of the system will simplify following the same line of thought. In the case where change in enthalpy is zero, T_0 becomes T_h;

(2.17)

. The change in enthalpy of the system simplifies as well

(2.18)

. The Gibbs free energy may now be written in simplified form as:

(2.19)

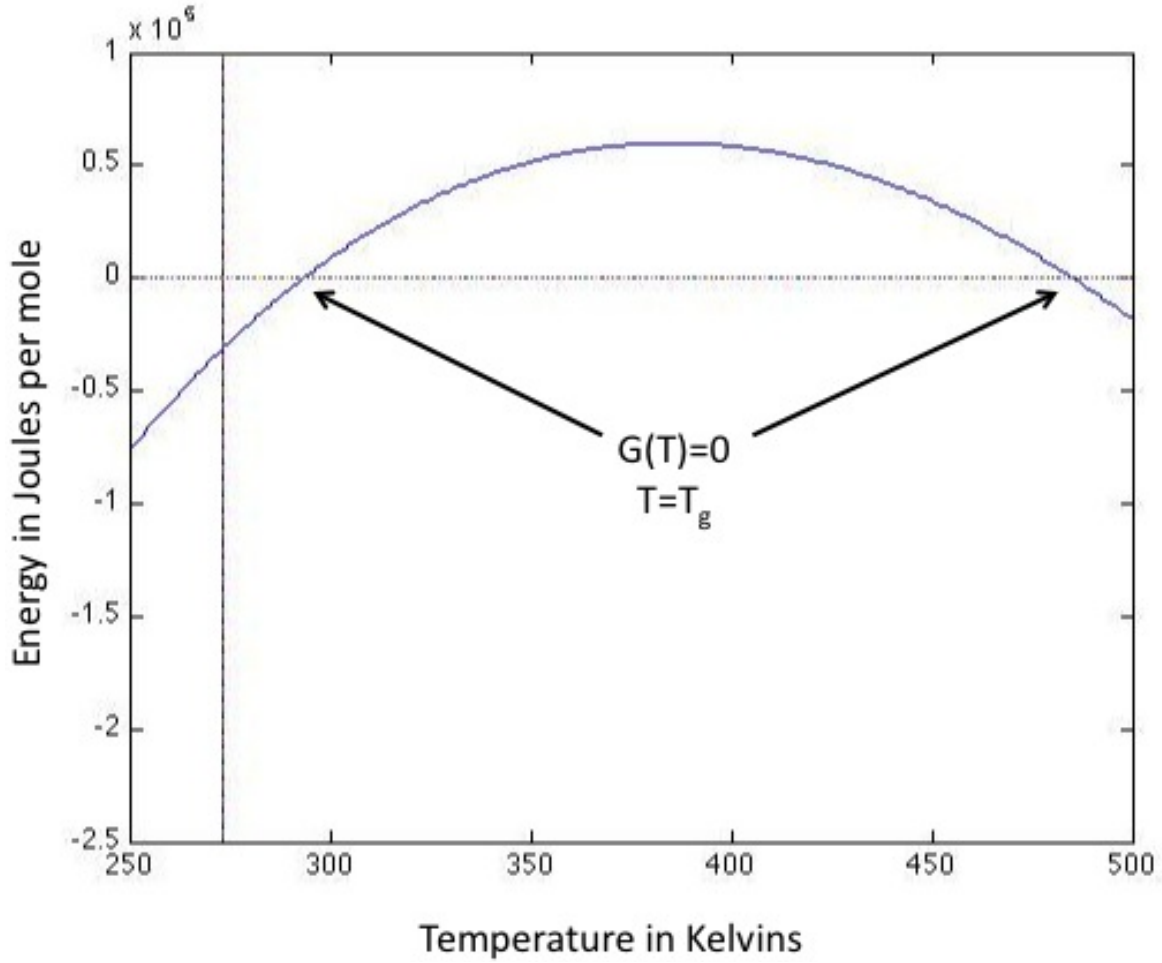


Figure 2.3: The gibbs free energy in Joules per mole graphed verse temperature in Kelvins.

. Figure 2.3 shows that when graphed verses temperature, the Gibbs free energy intersects the y-axis at a point where energy equals zero twice. This demonstrates that in two places, a colder temperature and a warmer temperature, the Gibbs free energy has no impact on the

likelihood of the system to change. At both of these temperatures, the system is independent of impact by the Gibbs free energy. Positive Gibbs free energy values are associated with stability. The system is most stable at the apex of the Gibbs free energy curve. This is where the protein should be found in a native state. Deviating in temperature from this apex suggests a tending towardinstability in the protein and potential folding. A protein will move toward instability and assume an unfolded conformation when there exists negative Gibbs free energy associated with the system.

Section 2.4: Defining the Temperature at Which the Change in the Gibbs Free Energy Vanishes

As the change in Gibbs free energy approaches zero , the variable temperature T_0 in equation (2.14) now becomes T_g. At this temperature, there is no usable free energy in the system without manipulation of its pressure or volume.

When the system approaches the temperature at which the change in Gibbs free energy is zero, the change in enthalpy and change in entropy must now incorporate the temperature limit. The enthalpy of the system is no longer guaranteed to be zero and T_g must be incorporated into the enthalpy equation

$$\Delta H(T) = \Delta H(T_g) + \Delta C_p(T - T_g) \tag{2.20}$$

. To incorporate the Gibbs free energy, it is no longer possible to ensure that entropy is zero. The system is moving toward disorganization at this point and entropy at the point where the Gibbs free energy is vanishes, is measureable

(2.21)

. Now that entropy and enthalpy have taken into consideration, the change in Gibbs free energy equal to zero can be taken into consideration. To do this, allow the variable T in Gibbs equation to move to T_g

(2.22)

.When the change in Gibbs free energy vanishes, the entropy and the enthalpy of the system scale by a factor of temperature

(2.23)

. In protein folding, T_g is the melting temperature of the protein or the temperature at which a phase change occurs. This phase change is the unfolding of a protein. With no other considerations made on the state of the system, Fig. 2.3 demonstrates a protein unfolding

temperature at 299K or 26 degrees Celsius, and 475K or 202 degrees Celsius. A protein in solution in a thermodynamic system thus has a tendency to unfold at both a cold and a warm temperature. Now, assume without support yet, that the heat capacity is constant throughout all further calculations.

The entropy and enthalpy of the system with respect to temperature can be rewritten taking into consideration the relation between entropy and enthalpy when the change in Gibbs free energy is zero

(2.24)

. The slope of the enthalpy of the system should be characterized by a negative entropy in cases where temperature is greater than T_s and to take this into account, the specific heat portion of the change of enthalpy approximation above is modified to directly represent this (Becktel 1987).

The entropy equation is also modified to take into account the desired slope for temperatures greater than the temperature at which entropy is zero

(2.25)

. Entropy can be approximated using the measured change in enthalpy and the temperature at which the Gibbs free energy vanishes.

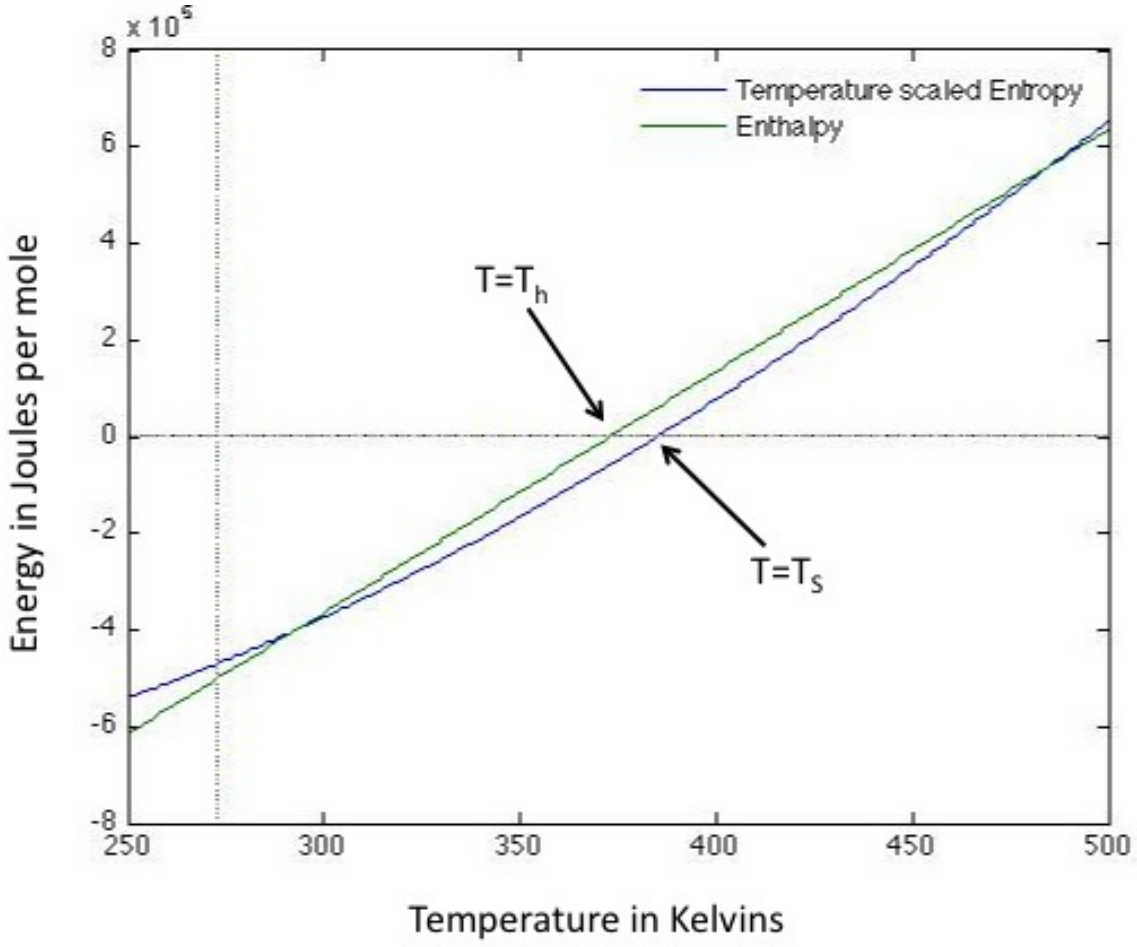


Figure 2.4: The components, temperature scaled entropy and enthalpy, of Gibbs free energy in Joules per mole graphed verse temperature in Kelvins. Where T_h is at 373 Kelvins and T_s is at 385 Kelvins are taken from the ideal thermal protein model for Cytochrome-C (Robertson 1997).

Figure 2.4 relates the scaled entropy of the system to the enthalpy of the system. Remember from Fig. (2.2) that the entropy increases to the right of T_s and change in the system is favorable. It can be inferred from Fig. (2.3) that the change discussed is the tendency of a protein to fold into its native state. Then here, to the right of T_s a protein will have the tendency to fold. From Fig. (2.1), T_g denotes the thermodynamic barrier created by a systematic independence from the heat of enthalpy. This means that, to the left of T_g, the system will favor protein folding into native confirmation. Figure 2.4 demonstrates the careful balance between the thermodynamics of the system that cause a protein to fold or unfold.

Section 2.5: Relating Variable Temperature Quantities

The temperatures described so far represent the thermodynamic characteristics of the system can be related directly. Through the incorporation of the limits of the native or folded conformational state to the limits of the denatured or unfolded conformational state, equations relating T_g, T_h and T_s can be formed.

Reapply the limit at which enthalpy vanishes

(2.26)

. Now substitute the approximation of change in enthalpy, equation (2.24), into the above relation to create a relation between T_g and T_h. Temperature, T, in the enthalpy approximation becomes T_h and the following relation arises:

(2.37)

. Now solving for T_h gives an equation that, knowing ΔC_p, $\Delta H(T_g)$ and T_g, will generate the temperature at which the heat of enthalpy is zero

(2.28)

. To find T_s, consider the entropy approximation, equation (2.25), and again apply the limit at which entropy is zero. This case will change the temperature T, of entropy approximation to T_s

(2.29)

. Now solve for the temperature at which entropy is zero:

(2.30)

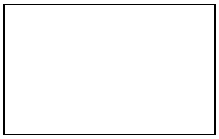
. The temperature at which entropy is zero is then,

(2.31)

. To eliminate the exponential denominator, an exponential expansion provides a convenient approximation (Robertson 1997) when the heat of enthalpy is much smaller than the specific heat multiplied by the temperature,

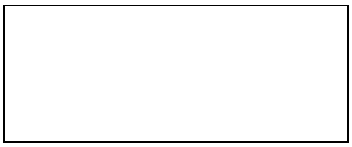
$$\Delta H << T_g \cdot \Delta C_p$$

then



(2.32)

. This can be simplified even more by incorporating the change in enthalpy with respect to T_g, equation (2.27),



(2.33)

. The temperatures expressing limiting characteristics of the system have now been related. The temperature at which entropy vanishes can now be written in terms of the temperatures at which the Gibbs free energy and enthalpy both vanish as well. The relationship between entropy, enthalpy and Gibbs free energy, as well as the temperatures at which they vanish, will later be used to further develop a model for folding.

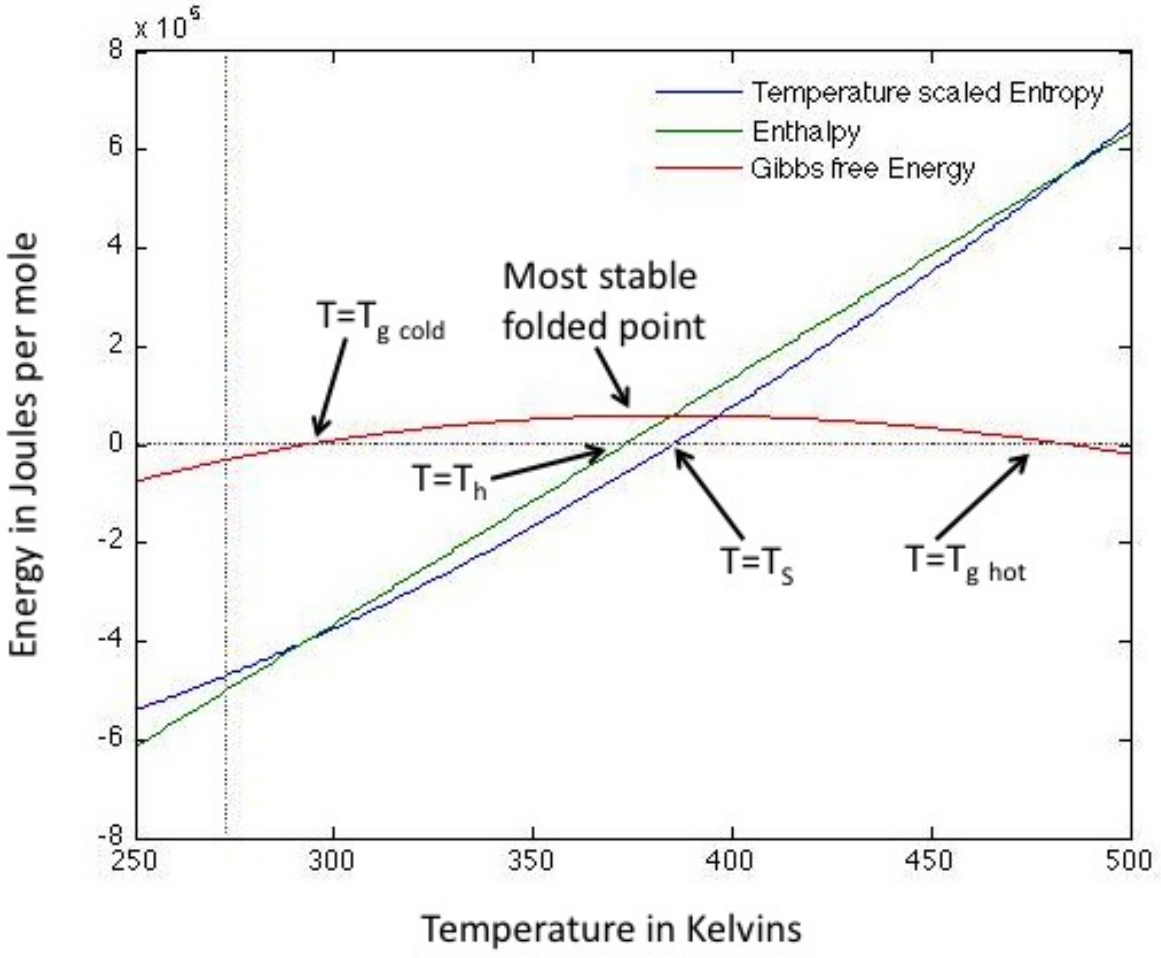


Figure 2.5: The components, temperature scaled entropy and enthalpy, of Gibbs free energy and the Gibbs free energy in Joules per mole graphed verse Temperature in Kelvins. Where T_h is at 373 Kelvins and T_s is at 385 Kelvins are taken from the ideal thermal protein model for Cytochrome-C (Robertson 1997).

A final comparison of the components of the Gibbs free energy as well as the Gibbs free folding energy itself will aid in understanding how the thermodynamics alone of the system affect a protein's tendency to fold. Figure 2.5 shows that the most stable point for a folded protein governed by the Gibbs free folding energy is also favorably governed by the Enthalpy because the apex of the Gibbs free energy curve is to the left of T_h. The apex of the Gibbs free energy curve falls in the arena of favorable folding due to enthalpy but not due to entropy. Because figures 2.1-5 model Cytochrome-C under ideal thermal conditions, the likelihood of thermodynamics being the only physical set of principles to dominate the system hinges on the entropy of the system. In the case of this thermodynamic modeling utilizing the ideal thermal protein, T_s suggests that entropy of this system is unfavorable to protein folding. The balance of the temperature scaled entropy and the enthalpy is vital to the role thermodynamics plays in the unfolding of a protein. The enthalpy of the system is the internal energy of the protein and increases with temperature making it easier to unfold as temperature increases. A system governed by enthalpy will tend to denature. If the system is governed by the entropy, disorder is

more likely as temperature increases and entropy must compensate for this for the system to thermodynamically unfold the protein. Thermodynamics is not the only set of principles that governs protein folding and other relations must further be considered to model denaturation.

Section 2.6: The Folding Free Energy

The Gibbs free energy can be directly related to the Helmholtz Free energy, F, of a system. The change in the Helmoltz free energy measures the amount of usable energy of a system when closed and at constant volume. The Helmholtz free energy here will be used to derive a relation for the free energy of a system of denatured and a system of native proteins. Because of its correlation to molecular folding of protein in the system, further references to the Helmholtz free energy will be known as folding free energy. The change in the Helmholtz energy of the system required to fold a protein will be denoted as the change in folding free energy required to move a protein from its denatured state to its native structure. Folding free energy will then be measured as the amount of free energy associated with the native state less the amount of free energy associated with the denatured state

(2.34)

. Notice that the Gibbs free energy and the Helmholtz free energy are both measurements of the usable energy of a closed system. Where they differ is the requirement of the system to be isobaric or isochoric. The relationship between these quantities is (Dill 2002):

(2.35)

. The product of pressure change, (Δp), and volume change, (Δv), over the range of our experiments can be estimated as

(2.36)

. Where all other values reported and used experimentally are several times on the order of magnitude of this pressure-volume change. For the accepted value of the specific heat of Cytochrome-C, 5.3 kJ K^{-1} mol^{-1} (Robertson 1997), the caloric energy to for a change in temperature of 1 degree Celsius is

(2.37)

. Where the energy is for one mole of particles. The energy to increase the temperature by one Kelvin is 3 orders of magnitude greater than the energy change due to change in volume and change in pressure. The energy needed to fold or unfold a protein is much greater than the change in pressure or volume. Therefore, we can approximate the Folding Free Energy as the Gibbs Free Energy:

(2.38)

. The folding free energy can now be written in terms of the thermodynamic conditions needed to derive the Gibbs free energy. Because the system in consideration works with macromolecules that are known to have charge distribution and are found in solution of diverse charge composition, a folding free energy associated with electrostatics will need to be addressed later in section 2.11.

The neutral folding free energy can be used to model folding of a protein after the enthalpy and entropy have been modified where again the system change is measured at static temperature,

(2.39)

. For large molecules, molecular packing affects the heat of enthalpy of the protein. Through experimentation, it has been found that the energy needed to bury an amino acid in a protein is directly related to the enthalpy of the system. For N amino acids in a protein chain, let the average energy required to bury an amino acid be denoted by g. Then the total average energy to bury amino acids in a protein can be calculated by multiplying the average energy by the number of amino acids in the chain. The change of enthalpy should now incorporate this burial energy. For an unfolded chain, the burial energy is non-existent. The burial energy is then directly associated with the change of enthalpy from the folded to unfolded state,

(2.40)

. At the temperature where enthalpy vanishes due to thermodynamic effects, the chain induced enthalpy remains,

(2.41)

. The energy required to bury an amino acid is the sole contributor to enthalpy at the temperature at which enthalpy does not change due to thermodynamic effects. The variable N in Eq. (2.41) is the number of amino acids in a protein. For Equine Cytochrome-C, there are 104 amino acids contained in the protein sequence (Privalov 1974). The variable g is a measured value dependent on the denaturant in solution as well as the concentration of the denaturant. Further explanation will follow in the next section.

Section 2.7: A General Overview of Structural Interaction Affecting Thermodynamic Enthalpy of the System

Several structural characteristics of proteins interact with surrounding solution and affect folding. Hydrophobic interactions with surrounding solution, hydrogen bonding between segments of the protein and Van-der-Waals interactions between solution and protein lead to changes in the derived thermodynamic equations to account for solution movement.

Recall that enthalpy measures the tendency of a system to change its internal energy. Because solution-protein interactions affect the folding of a protein, any changes in solution conditions must be accounted for. Adding denaturant to solution will affect the ability of a protein to fold. To account for this relationship to folding, the molarity of the denaturant is multiplied by a calculated value for each denaturant. Guanidine HCl is the denaturant used to cause unfolding of protein in these experiments. The effective quantity that relates the concentration of denaturant to the energy of unfolding of a system is m_1, given by (Dill 2009) as 25 for Guanidine HCl. As denaturant is added to a protein solution, the hydrogen bonds on the exterior of the protein weaken and the protein begins to unfold.

The concentration of the denaturant in solution is represented as c, in units of molarity,

(2.42)

. The change in heat of enthalpy of saturation is the enthalpy that is directly affecting protein folding due to solution concentration. This can then be incorporated with the average energy per amino acid after an intrinsic average energy per amino acid, g_o, is established. The intrinsic average can be measured or approximated using an ideal thermal protein (Dill 2009). The ideal thermal protein is a construct of protein characteristics that demonstrate the structural influence to the thermodynamic aspects of the system. The model creates a theoretical protein that will unfold in favorable conditions and fold under unfavorable conditions. The model also incorporates the theory that cold denaturation exists in proteins under the right conditions. To date, there is no measured intrinsic average energy per amino acid values for the protein used throughout the experiment. The ideal thermal protein uses an average packing energy of -1200 cal/mol(Dill 2009). The average packing energy was calculated through experimental calorimetric testing, as

(2.43)

. The average energy for protein packing (2.41) can be modified to incorporate the change in enthalpy saturation (2.42) and the intrinsic average energy of packing

(2.44)

. The total change length can easily be reincorporated by multiplying through the total average energy of packing per amino acid considering the molarity of the of denaturant in solution (2.44) by multiplying through by chain length N as

(2.45)

. The average energy for protein packing, Eq. (2.43), multiplied by the chain length of the protein is always negative. This enthalpic term always works to decrease total enthalpy and, like entropy, will act to encourage bond formation or order in the system. Negative enthalpy encourages folding of the protein. The denaturant concentration multiplied by its intrinsic value, represented as m_1, and the chain length of the protein is always positive and works to increase the enthalpy. An increase in enthalpy works to disorder bonds and unfold the protein. The balance between these opposing terms will prove important in modeling protein folding.

The enthalpy due to the effects of the denaturant in solution and the interaction between the protein and surrounding solution can now be incorporated into the thermodynamically derived enthalpy as

(2.46)

. The limit at which the thermodynamic affects of the system are no longer a determinant of the system's behavior again causes this new value of enthalpy to reduce to the limit taken in relation (2.41):

(2.47)

. At $T_{h'}$ the enthalpy of the system is dependent on denaturant concentration, a temperature at where thermodynamic enthalpy no longer affects the system. Incorporating the structural characteristics into the thermodynamic enthalpy of the system allow for a stronger prediction of protein behavior in solution. Recall that at T_h, there is no enthalpic contribution to the system. Because of this, Eq. (2.47) must go to zero

(2.48)

. Substituting the definition for g into Eq. (2.48) will demonstrate the necessity of each component value,

(2.49)

. Unless the number of amino acids in the protein is zero, the packing energy g_o must be negative to count act the positive molarity measured in an experiment and the effective quantity that relates the concentration of denaturant to the energy of unfolding of a system. When the change in enthalpy at or near T_h does not go to zero, it is incorporated into the total change in entropy and will affect the folding free energy.

Section 2.8: A General Overview of Structural Interaction Affecting Thermodynamic Entropy of the System

Entropy must be modified to incorporate both structural and thermal entropy. The reduced entropy of a system (2.16) represents thermally developed entropy. The structural attribution to entropy is given as (Dill 2009);

(2.50)

. Where N is the number of amino acids in the protein; for Cytochrome-C, N is 104, k is Boltzman's constant and z is the number of rotational isomers possible around the carbon backbone per amino acid of the protein. No reported values of the possible number of rotational isomers have been reported for Cytochrome C and for further calculations, the ideal thermal protein value will be used as follows

(2.51)

(Dill 2009). For the ideal thermal protein, the possible number of rotational isomers is an average and does not correspond to direct physical interpretation. Again, entropy due to structure is considered the only contributing factor at temperatures where the entropy due to thermodynamics vanishes. The structural denatured entropy is always positive and must be overcome to unfold a protein. The energy needed to overcome this entropy component increases as entropy increases as a function of temperature.

The total change in entropy is a sum of all possible changes in entropy

(2.52)

. Using the simplified definition of entropy (2.23) as the thermally attributed entropy, the total change in entropy is

(2.53)

. Entropy has now been written in terms of a structural and thermal contribution. Taking into both the system, it's contents and surroundings allow for a closer prediction to the actual behavioral model of a protein in solution.

Section 2.9: The Neutral Folding Free Energy

The neutral folding free energy can be formed by incorporating the newly defined entropy and enthalpy that now consider both the structure and thermodynamics of the system. Substituting in the newly developed enthalpy (2.46) and entropy from (2.53) into the folding free energy equation (2.34) give the final relation for the neutral folding free energy

(2.54)

. The neutral folding free energy of the system is due to uncharged characteristics of the solution and system. The entropy and enthalpy components of the system, intrinsic to the protein in consideration determine whether the neutral folding free energy is enough to unfold the protein. In the case of Cytochrome-C, the thermodynamically derived components prevent naturally occurring unfolding of the protein. Figure 2.6 shows that protein specific, chemically derived thermodynamic components are able to overcome the thermodynamic components for Cytochrome-C and induce unfolding. Unfolding occurs when the folding free energy is equal to zero. The neutral folding free energy is only effective enough to induce unfolding as extreme temperatures. If the solution has not frozen, unfolding will occur well below freezing and at over 900 degrees above boiling. Both hot and cold denaturations are possible when the neutral free energy is considered but are hard and unlikely to replicate in laboratory setting. Temperatures below – 50 degrees Celsius are difficult to produce in an exposed area or outside of a well insulated – 80 degree Celsius freezer. The sample mount and small angle x-ray scattering setup that will be discussed later, would be impossible to be contained in an insolated atmosphere below – 50 degrees Celsius. Experimentation discussed in chapters 4 and 5 demonstrate that temperatures below – 25 degrees Celsius were difficult to obtain.

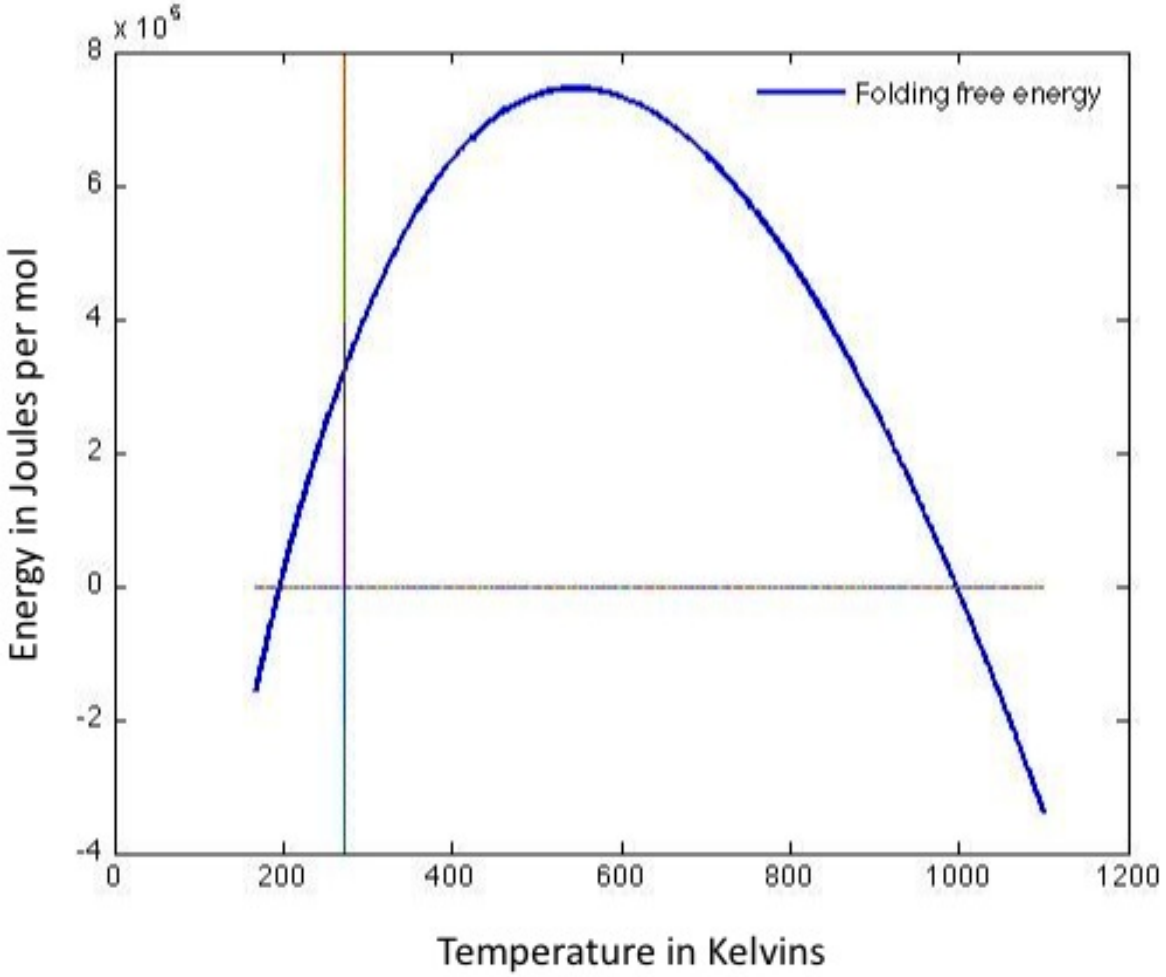


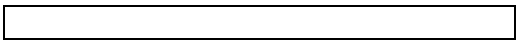
Figure 2.6: The Neutral Folding Free Energy for 2 molar denaturant using Gibbs free energy values from figures 2.1-2.5, the z value from eq. (2.59), N is 104 amino acids for Cytochrome-C and Boltzmann's Constant (k = 8.31 Joule/mol/K), C_p for Cytochrome-C, 5000 Joules/mol K, g_o for the ideal thermal protein, -1200, as well as, c, the concentration of denaturant, 2, and m_1= 25 cal/mol for the denaturant Guanidine HCl.

In a naturally occurring solution, a system will have charged characteristics that must be accounted for before the folding free energy of the system can correctly model the a system of interest.

Section 2.10: The Electrostatic Folding Free Energy

The electric folding free energy is the folding free energy that accounts for the free energy of the system causing protein folding due to electric charge interaction between the protein and the surrounding solution. The chemical structure of macromolecules and, thus proteins affect the charge distribution throughout the solution as well. These charges, together, are needed to develop a relation for the folding free energy of the system.

The total change in the folding free energy must be written as a sum of both the change in the neutrally developed folding free energy and the change in the folding free energy that is dependent on chemical and electrostatic conditions



(2.55)

. From this equation, it becomes immediately important to guarantee that the neutral folding free energy and the electrostatic folding free energy are on the same scale. The Bjurrem length, l_b, is the distance at which electrostatic interactions between charges are on a comparable scale to the thermal energy of a system. The Coulomb energy will be approximately equal to the thermal energy at this length (Dill 2002). To determine this length, allow the distance in the coulomb energy to be represented by l_b. Setting the Coulomb energy equal to the thermal energy and solving for distance defines the Bjurrem length:

(2.56)

.Where ε is the dielectric constant of the surrounding solution, ε_o is the permittivity of free space and e is the charge of the electron. Again, k is Boltzmann's constant. Solving for length to get Bjurrem length

(2.57)

. Numerically, the Bjurrem length may be written as:

(2.58)

. The constant value C absorbs the numerical difference between boltzmann's constant, k, and the ideal gas constant, R, as well as the value 4π. The Bjurrem length now relates the strength of the folding free energy due to neutral contributions to folding free energy due to the electrostatic effects of the system discussed below in section 2.11.

Section 2.11: Building the Electric Free Folding Energy Model Using Basic Electrodynamics

To model the electrostatic folding free energy for proteins, the surroundings of the protein must be considered. Salts in solution dissociate once placed in water and behave as positive and negative charges. Small proteins have localized charges due to the varying side

chains on individual amino acids. The dissociated ions dissolved in solution surround these small proteins and act to shield the macromolecule from exterior potential fields.

First assume that the small protein in solution is approximately spheroid with a surface charge, *Q*, in salt solution, the sphere is shielded by ions, reducing the potential strength of the sphere outside of the ionic barrier. Let $\Phi(x)$ be the electrostatic potential at *x* away from a planar surface. A planar surface is chosen because distances far away from the surface will be considered. The concentration of mobile positive ions in solution can be represented by $n(x)_+$ and the concentration of mobile negative ions in solution is represented by $n(x)_-$. Using the Nernst Boltzmann equation, both the concentrations of negative ions and positive ions can be modeled as

(2.59)

(2.60)

. Here, *z* is the number of electric charges and is the concentration of ions in the total solution. The total charge density, ρ(x), can be written as

(2.61)

.Where z_i represents the valence ions of species *i,* all at a distance *x*.

The Poisson Equation relates the charge density to the electric potential,

(2.62)

. Substituting Eq. (2.61) for the density into Poisson's Equation yields

(2.63)

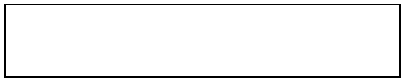
or,

(2.64)

. Inserting the expressions for the salt concentrations gives

(2.65)

. Equation (2.65) looks much like a trigonometric function. One possible function to consider is the hyperbolic sine function

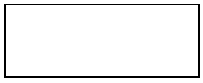


(2.66)

. Poisson's equation can be rewritten

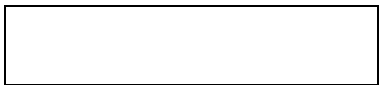
(2.67)

. If the electrostatic potential is small, or,



(2.68)

, then the small angle approximation can be used to simplify the hyperbolic sine function as



(2.69)

1

. Under this approximation, Poisson's equation can be written as

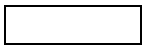
(2.70)

. Here, the parameter κ, called the linearized Poisson-Boltzmann constant, has been introduced to simplify the equation

(2.71)

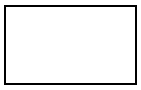
. Solving for κ^{-1} gives the Debye length. The Debye length is the length at which the shielding of ions affects the potential.

The salt concentration, c_s, of a solution can be represented by the population of ions in solution multiplied by the valence charges in solution (Dill 2005)



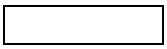
(2.72)

. Following literature derivation (Dill 2005,Dill 2002), the Bjurrem length can now be written approximately



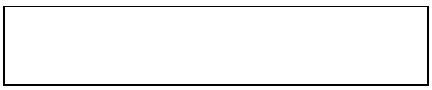
(2.73)

. Where, from eqn. (2.58), the scaling constant, C can be ignored and the Bjurrem length may be approximated. Using both the formula for the salt concentration and the Bjurrem length, the linearized poisson-Boltzmann constant can be written as



(2.74)

. If a small protein is modeled as a sphere of radius *a*, on its surface there will be total charge *Q*. The electric potential is directed radially outward. The potential modeled above can be used where the distance measured is the radial distance from the center of the sphere. Poisson's equation can be rewritten in terms of this radial distance. Further, because the electric potential is only in the radial direction, Poisson's equation is as follows



(2.75)

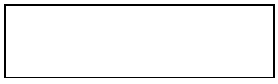
. Rearranging the equation gives a second order differential equation

(2.76)

. The most general solution of this equation is

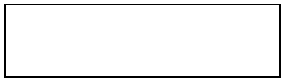
(2.77)

. Where the constants A_1 and A_2 are found by applying boundary conditions. The electric potential must go to zero at infinity. To satisfy this, A_1 must be zero, which leaves



(2.78)

. The surface of the sphere located at a radius a, will now be used to find A_2

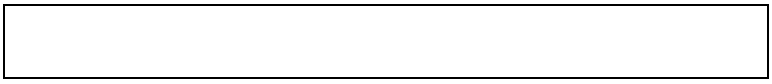


(2.79)

. Solving for A_2 gives

(2.80)

. Knowing A_2 allows the electric potential to be written as



(2.81)

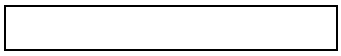
. To find the surface potential at a radial distance, a, derivative of equation (2.81) is taken with respect to radial distance and considered at the boundary

(2.82)

. Applying the limit where the radial distance becomes the radius of the circle, the partial derivative of the electric potential with respect to the radial distance can be rewritten as

(2.83)

. Using



(2.84)

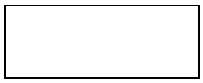
, the partial derivative simplifies to

(2.85)

or,

(2.86)

. The derivative of the electric potential is the negative electric field



(2.87)

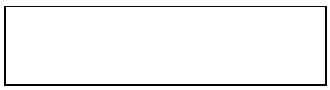
. The electric field for a sphere has the general form:

(2.88)

. At the surface boundary of the sphere, the electric field is a function of the sphere's radius, a,

(2.89)

. This can be used to solve for the electric potential when both equations (2.86) and (2.89) are inserted into equation (2.87)

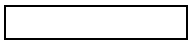


(2.90)

. The electric potential on the surface of the sphere is then

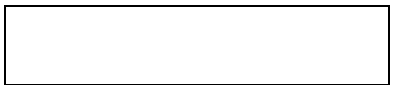
(2.91)

. Electric potential is directly related to the potential energy of a system, U, by multiplying through by charge. The potential energy of a system is then directly related to the electrostatic contribution of the folding free energy, F_{elec}



(2.92)

. In this case, the folding free energy at the surface of a sphere with radius, a, is then:



(2.93)

. The Bjurrem length can be rearranged such that it conveniently fits into equation (2.93)

(2.94)

. Now use this rearrangement to rewrite the folding free energy

(2.95)

. Let the total number of charges in the system be defined as

(2.96)

. Where Q is now the net number of charges, due to the pH of the solution and the pH of the protein being modeled as the sphere. The net charge can also be defined as a difference between the basic and acidic net charges on a given protein due to its specific reactive backbone

(2.97)

. Proteins contain both acidic and basic amino acids. Q_b is defined, (Dill 2009) to be a sum of pH relations of the basic amino acids in the protein. The pk_i is the proton dissociation constant of a given amino acid and b represents the number of amino acids with volatile pH, Q_b is given to be:

(2.98)

. The acidic net charge for , a, the number of reactive side chains is

(2.99)

. The inversion of pH and pk_i comes from the relation of pk_a to pk_b. Values of pk_a and pk_b For Cytochrome-C's reactive groups can be found in Appendix 2 and Section 5.5.1.

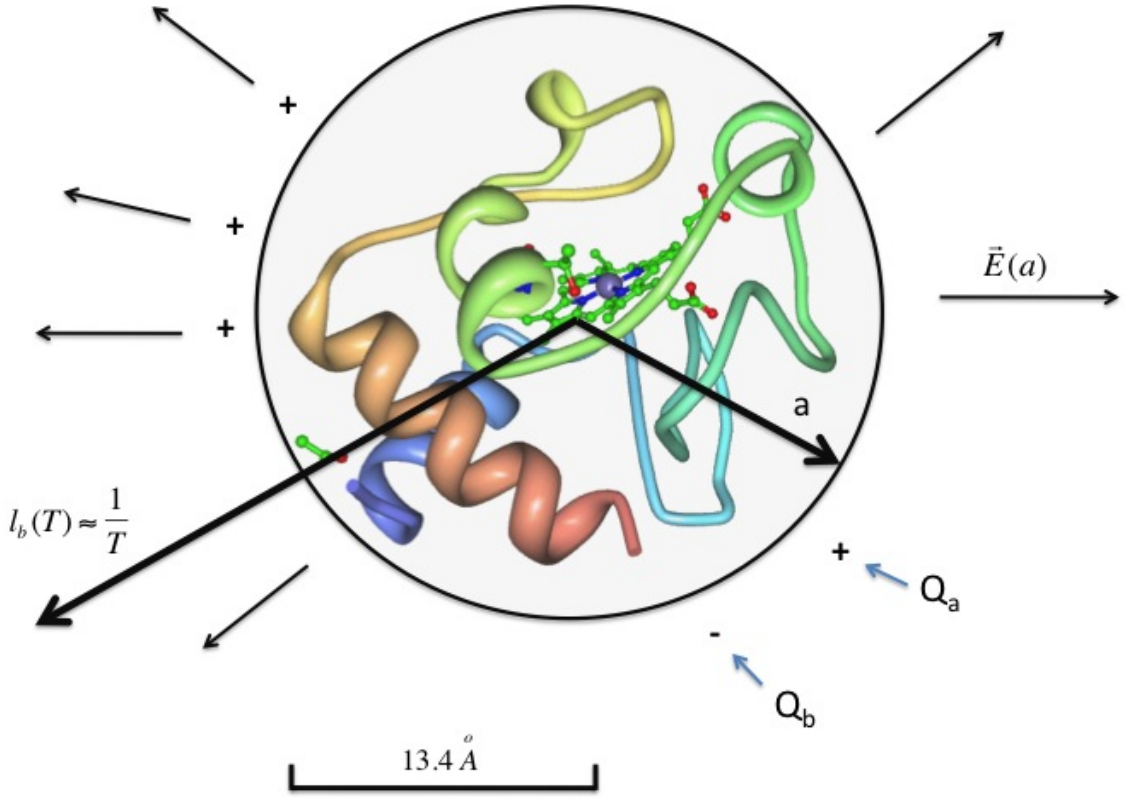


Figure 2.7: Relating the Bjurrem length as a function of inverse temperature to the Electric field of the protein where the protein is modeled as a sphere of radius a, with charges on its surface.

Figure 2.7 directly demonstrates the relation between the temperature dependent, Bjurrem length and the Electric field of the protein. The Electric field is localized due to structural charge distribution intrinsic to the protein. Because of this, the negative and positive charges are taken into account separately.

Now that Q is known as well as the Bjurrem length and the Debye length have been defined, the change in folding free energy can be defined. The change in energy that will cause a protein to fold is the difference between the folding free energy of the native and denatured states

(2.100)

. Equation (2.95) in combination with equation (2.96) can now be defined as the electrostatic folding free energy and rewritten with respect to its dependent variables

(2.101)

. At the point of denaturation, the radius of gyration of the protein is R_d. The radius of gyration takes the place of the radius of the sphere in equation (2.95) to give the denatured folding free energy due to electrostatics

(2.102)

. The net charge of the system changes as the protein folds and unfolds and must be calculated for both the denatured and native states. This necessity is demonstrated by the new subscripted Q_d.

At the point of complete folding, the native state has a radius of gyration represented by R_n. Using the radius of gyration for the folded protein in the native state will generate the folding free energy of the protein for electrostatic conditions in the native state

(2.103)

. Both equations (2.101) and (2.102) can be combined into equation (2.100) to get the change in folding free energy of a system due to electrostatic interactions

(2.104)

. Finally, using equation (2.42), the total change in folding free energy is given to be a combination of the electrostatic and neutral folding free energy

(2.105)

. The folding free energy now can be shown to correctly model protein folding due to several structural, solution based, electrostatic and thermal contributions. As will be shown later, the consideration of the electric folding free energy in modeling unfolding of a protein makes hot and cold protein denaturations accessible in the laboratory setting.

Chapter 3

Small Angle X-ray Scattering

Section 3.1: Overview

This chapter outlines the theory of small angle x-ray scattering (SAXS) as applied to proteins in solution at low concentration. Starting with the calculation of scattering from a single electron, the theory is extended to more complicated systems by determining the phase difference between different scattered waves illuminated by a coherent source. This result is generalized to a continuous charge distribution where the total diffraction intensity is given by a Fourier transform. Finally, we consider the special case of spherical particles and derive the low-angle dependence of the scattered intensity on particle radius (the Guinier Approximation), which is used to analyze the data in this thesis.

Section 3.2: Introduction to X-rays

The term X-ray defines a large range of wavelengths outside of the visible spectrum of light. X-rays can be characterized by wavelengths between 0.5 and 2.5 Angstroms. X-rays interact with matter by scattering and absorption.

X-rays can either elastically or inelastically scatter off electrons. Elastic scattering occurs when the photon loses no energy during interaction. Energy is conserved in the interaction and only the direction of propagation changes. Inelastic scattering, or Compton scattering, occurs when the incident photon loses energy and changes direction upon scattering off of a surface. The photon loses energy causing an increase in the wavelength.

The photoelectric effect occurs when x-rays are absorbed by matter and may be re-emitted. The reemitted wavelengths are directly dependent on the chemical composition of the sample absorbing the photon. The photoelectric effect will be ignored throughout this experiment.

Section 3.3: Defining the Scattering Angle

X-ray scattering can be described by the angular deviation of the x-ray caused by a scattering center. To demonstrate this relation, allow an x-ray to propagate along the x-axis toward the origin of a coordinate system. An electron resides at the origin and is denoted as point *O* in Fig. 3.1. The incident x-ray is represented by the unit vector $\hat{}$ and is normalized by the wavelength of the x-ray. As the x-ray nears the electron, it elastically scatters off the particle's surface and is deviated from its original trajectory by an angle of 2θ. The scattered x-ray is denoted by the vector, $\hat{}$ and is also normalized by the wavelength of the x-ray. The vector, $\hat{}$, relates the scattering and incident x-ray vectors as well as the wavelength

(3.1)

. Bisecting the angle 2θ and the vector , at it's midpoint, *P*, with a line segment, **OP**, provides the right triangle needed to express in terms of a trigonometric function. From Fig. 3.1, it is clear that is also directly related to the angle θ

(3.2)

. Because the incident and scattered x-ray vectors are defined as unit vectors, can be expressed as a direct relation between the wavelength of the x-ray and half the scattering angle

(3.3)

. This relation for the scattering angle will be used in further calculations to eliminate an overabundance of variables and simplify theoretical formulation.

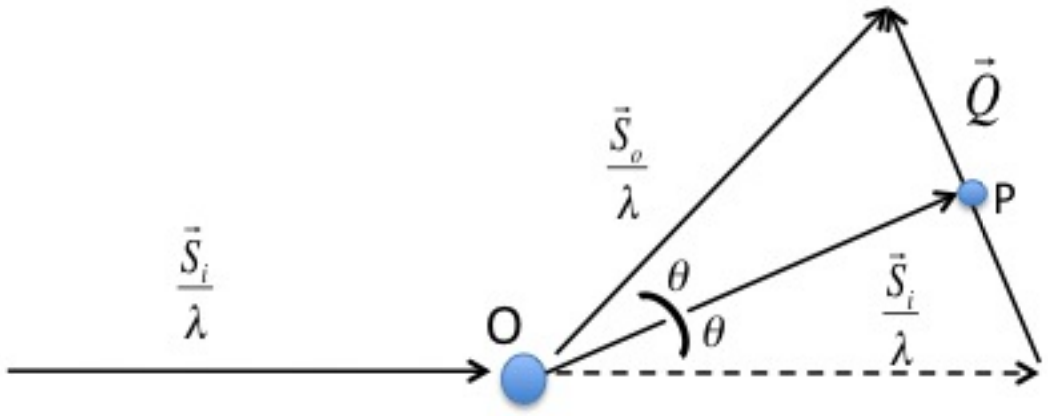


Figure 3.1: Defining variable vector with respect to incident and scattered x-ray vectors and scattering angle 2θ.

For every scattering center, there is a corresponding angular deviation in the intensity of the incident ray and a vector, , that describes it. In the case of multiple scattering centers experiencing elastic scattering, interference arises in the scattered intensities. Considering more than one scattering center at a time and looking at the individual scattered rays will make the occurrence of interference more clear.

Section 3.4: Examining Phase Shift through Interference Development

Considering the elastic interactions of a simple system of scattering centers will demonstrate how scattering waves form a distinctive interference pattern. The simplest case study of a multiple scatterer system is a two electron system. Allow one of the electrons to sit at the origin of the system and a second electron to sit a distance x away. The location of the second electron is described by the vector □. If two parallel, identical x-rays are incident on both, identical, scattering centers, identical scattered x-rays should emerge from the elastic interactions. Because these electrons are located a distance apart, the scattered waves may or may not be in phase.

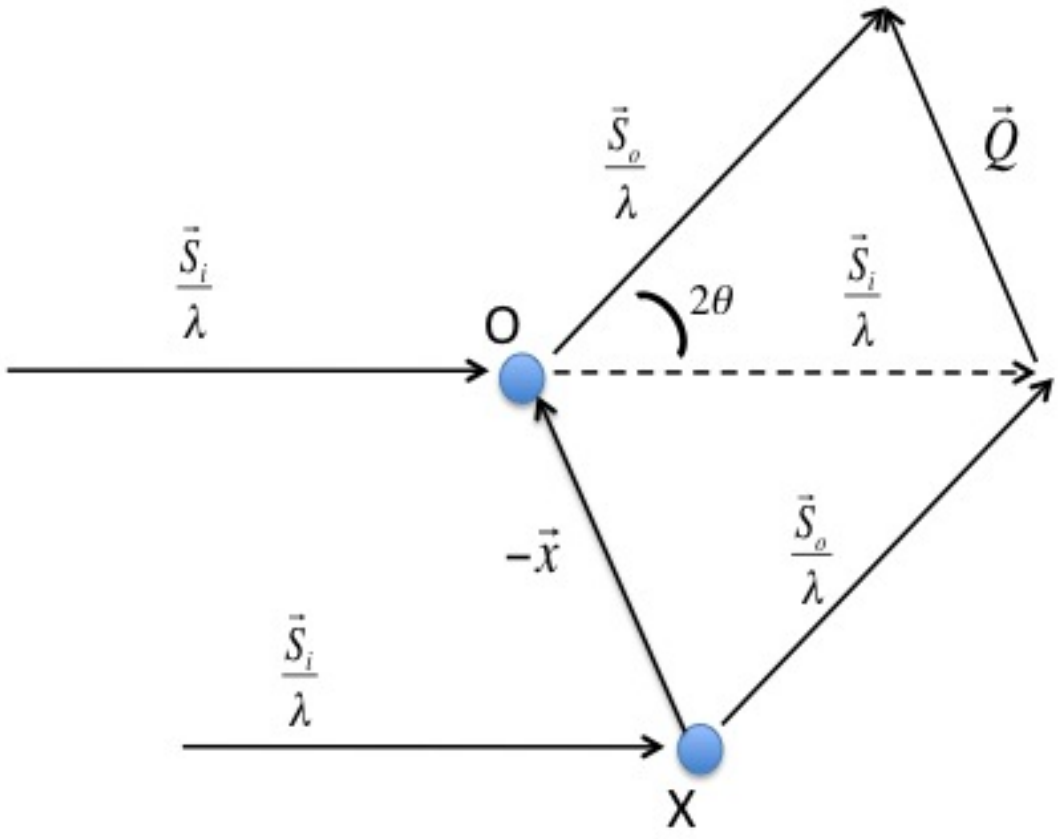


Figure 3.2: Defining □, the location of a second scattering center in the two scattering center system.

Notice that Fig 3.2 implements a basic property of vectors, directional inversion. The vector □ is represented by -□ so that it may point in the same general direction of □. Taking the dot product of the vector, □ and □ measures the projection of □, the vector describing the angle of scattering, onto the vector description of the position of the second scattering center.

The phase difference measures how much the amplitudes of both scattered waves align. Using the dot product, the phase difference, φ, can be defined by the projection of the vector description of the scattering angle, □, onto the vector description of the location of the scattering centers

□ (3.4)

. Again, the phase difference incorporates a negative from inversion of the position vector □. Now that the phase shift has been defined, an amplitude in terms of this shift in phase can be written.

The amplitude, A, of an x-ray scattered from an electron can be written in terms of the phase shifter using Euler angles

$$A = A_{\circ} e^{i\varphi} \tag{3.5}$$

. In the case where multiple scattering centers occur, the electron density of the scattering center will affect the amplitude, A. For an electron density of ρ, the amplitude of an x-ray off of a scattering center with multiple electron amplitude

□ (3.6)

. If the assumption is now made that the interference patterns of several x-rays off of several scattering centers add, the amplitude becomes a summation of amplitudes

□ (3.7)

. This assumption holds for all x-ray scattering unless scattering is considered in the very near field of observation. In the case of small angle x-ray scattering, the far field is considered and the associated interference patterns in this area correspond to the phase shift of the x-rays. Further, for a scattering center, or a sum of scattering centers with inhomogeneous electron density, the density can be written as a function of the location of the scattering center □ away

□ (3.8)

. The next step is to move from a summation of scattering centers to a continuous system of scattering centers. Over a continuous space, the amplitude can be defined in terms of the dot product of the distance between scattering centers, □, and the scattering angle, □

□ (3.9)

. Integrating over the volume will allow consideration of amplitude contributions in all three dimensions.

Now that we have the amplitude for an x-ray off a scattering center, the intensity can be calculated. Intensity is defined as the modulus squared of the amplitude (Guinier 1994)

(3.10)

. Using this relationship, the magnitude of the amplitude, A_o, is required to calculate the intensity. To eliminate this variable, the Thomson formula which describes scattering from an isolated election and can be used (Guinier 1994)

(3.11)

. For an incident beam with power measured in intensity per meter squared, I_e is the scattered energy for a unit solid angle per second. The classical radius of the electron is known to be

(3.12)

.To eliminate variables, the ratio of intensities can be used to write a ratio of amplitudes and thus generate a scattering relationship between electron charge density and scattering angle

(3.13)

. Separating the integral, the absolute value can be redefined as a function of

(3.14)

. Further, in the case of amorphous or small crystals with inhomogeneous density, the density function becomes a linear combination of the sample density and the electron density in the sample

(3.15)

.The density, ρ, is that of the protein sample and the density, ρ_o, is that the electron population of the protein sample. The resultant sum gives the density due to everything but the electron scattering centers. For proteins, this density is substantial. These changes can be substituted into the intensity ratio to give an intensity relation to the vector

(3.16)

. Assuming the sample is homogeneous, the density is constant, real, and can be pulled out of the absolute value

(3.17)

.This is the scattering intensity of the diffracted amplitude dependent on both sample density and the scattering vector .

The scattering intensity and the scattering angle is now a globally applicable relationship. It is now possible to consider the special case of the radially symmetric sphere in polar coordinates. Recall that the function looks much like a Fourier transform. The associated form factors, , vary inside and outside of the sphere in consideration. Inside the sphere, the form function is unity

, (3.18)

. Outside the sphere of consideration, there exists no sample to scatter off of and therefore the form function is zero

, (3.19)

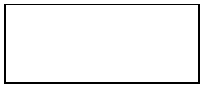
. Inserting the form function of the inside of the sphere into the Fourier transform relating and and integrating over all space will generate the form function, , needed to find the intensity

(3.20)

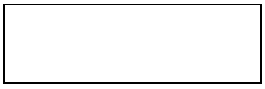
. Using the Fourier transform published for the radially symmetric sphere (Guinier 1994) and the associated form factor for the inside of the sphere, the form function, , can be expressed as

(3.21)

. For a full derivation of the Fourier transform of (3.20) please see appendix 3. Before the scattering intensity is calculated, there are a few assumptions that may be made to reduce the form function. In the case of this experiment, x-ray scattering patterns were observed at a distance of over a meter away from the sample. The scattering patterns observed were of radial distances smaller than .1 meters. Angles in consideration for small angle scattering were never greater than .002 degrees or .002 radians. This falls into the category of small angle approximation. Recall that both sine and cosine can be approximated for small angles

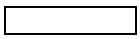


(3.22)



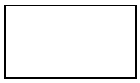
(3.23)

. For reasons that will be seen shortly, it is of necessity to expand the cosine term out to the third order. The form function can then be reduced for small angles. To simplify the calculation briefly, the following angular substitution can be made



(3.24)

. Next, the function can be simplified further by taking note that the volume of a sphere of radius *R* appears in the form function. Substituting the definition of volume into the function can further reduce the complexity of the form function

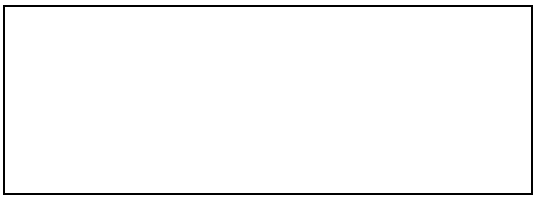


(3.25)

. Both reductions simplify the form function:

(3.26)

. Now the small angle approximations can be applied to further reduce the form function. To account for the angular simplification, the function shall be called by the change in variable to *u* divided be the terms independent of the form function

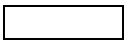


(3.27)

. Algebraic manipulation simplifies the form function into to a more workable form

(3.28)

. Recall that an exponent can be approximated to a Taylor expansion



(3.29)

. The form function can now be expressed as an exponent in the case of small angles

(3.30)

. It is now appropriated to reintroduce the definition of *u* into the form function and continue with the derivation of the scattered intensity as a function of 

(3.31)

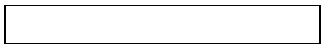
. From equation (3.17), the relationship between scattering intensity and the form function was defined with respect to the square of the form function as well as the square of the density function. Substituting the exponential form of the form function into the equation for scattered intensity will give a reduced form of the intensity function

(3.32)

. Notice, that the density multiplied by the volume results in the number density of x-ray scatterers per volume, *n*

(3.33)

. Incorporating the number density simplifies eq. (3.32) into a working function of the scattering intensity can be defined in terms of *n,*  , and 



(3.34)

.This form is much simpler than one may have expected but when graphed, the validity of the small angle approximation is made clear.

1

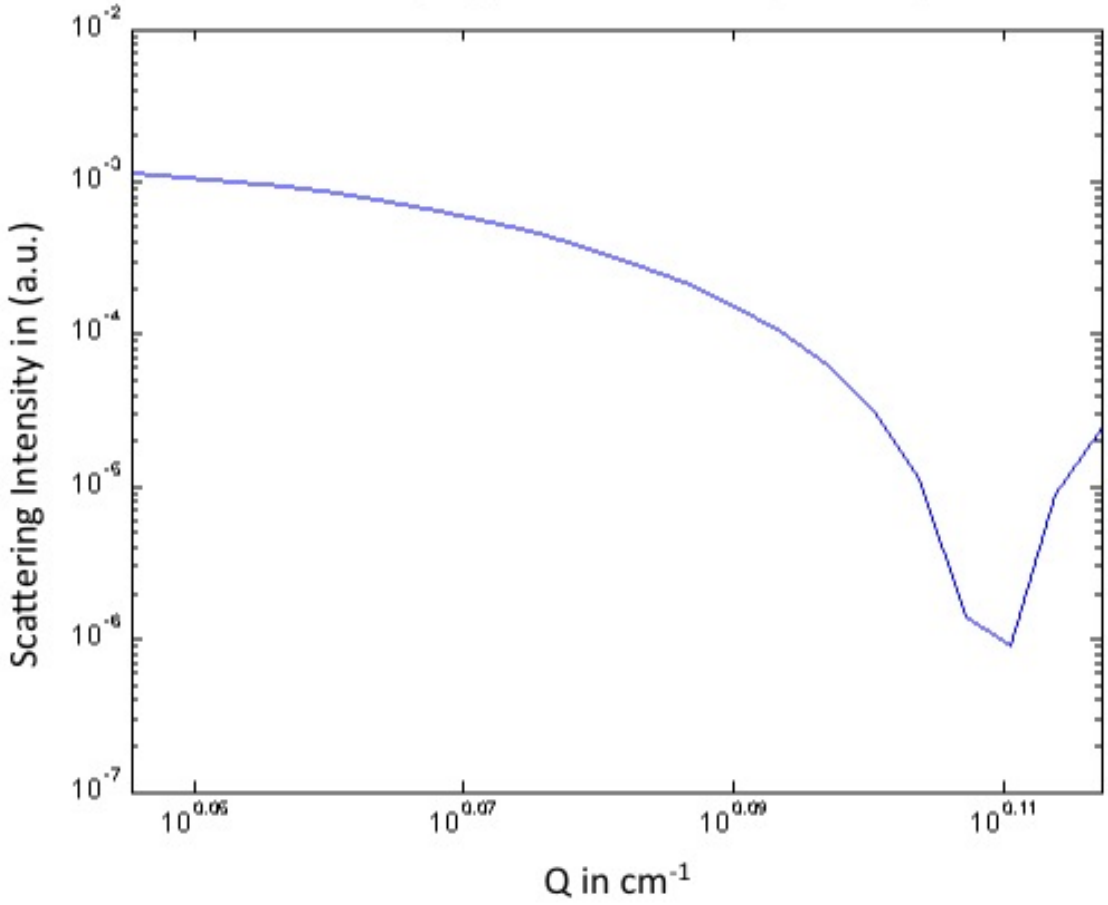


Figure 3.3: Theoretical plot of the scattered intensity as a function of the azimuthally symmetric radial distance in inverse centimeters. The radius of gyration used for the plot was 1.35 nanometers. This is the standard shape of an azimuthally averaged Intensity curve on a loglog plot.

In the laboratory, the intensity is measured directly using charge-coupled device, (CCD). The distance *Q* directly correlates the intensity at any location in the scattering pattern. Solving for *R* or the radius of gyration is accomplished by removing the variables from the exponent and utilizing the scattered intensity at a radial distance □ from the beam center.

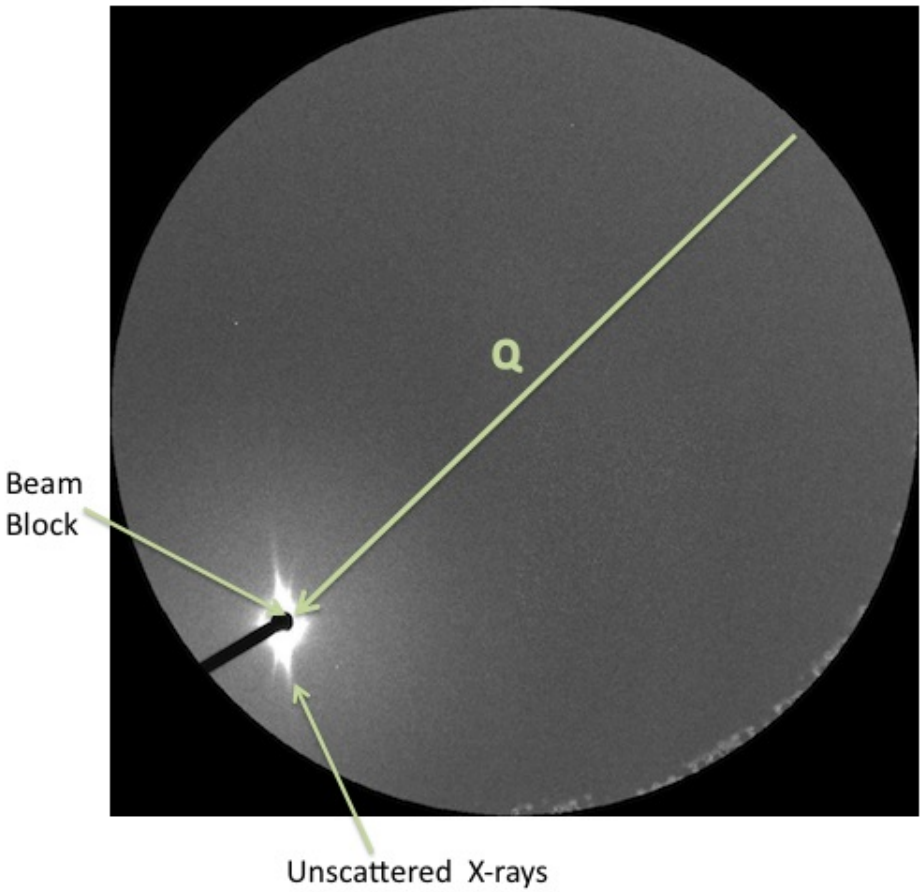


Figure 3.4: Actual unprocessed data as it is acquired by the CCD. A beam block protects the CCD from damage from unabsorbed x-rays in the direct beam. X-rays may still pass through the sample with only minor deviation; this occurrence shows up as a bright, diamond shaped, spot around the beam block.

Using the data from Fig. 3.4 will allow for the following theoretical calculations to reveal the radius of gyration of the protein. Solving for the vector ***R***, or the sample protein's radius of gyration begins by taking the natural log of both side of equation (3.34)

(3.35)

Finally, a fully reduced form of the scattered intensity measured in lab dependent on the radial distance from the beam center can be expressed.

(3.36)

Where the natural log of the number density squared is the y intercept of a Guinier plot. The simplest way to find the radius of gyration of the protein, R, is to graph the log intensity verse the $-2\pi Q^2$ and approximate a linear fit to the graph. The resultant slope would be R^2.

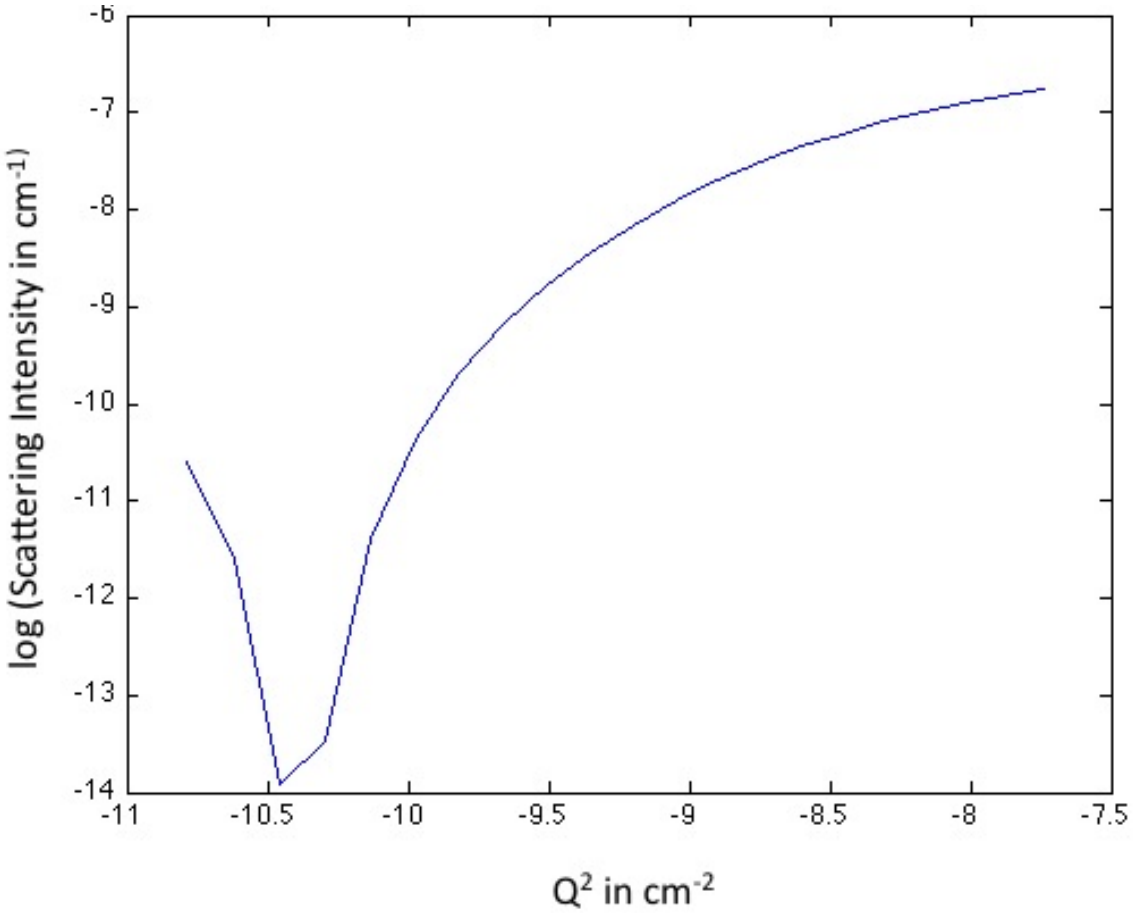


Figure 3.5: Theoretical plot of the natural log of the scattered intensity as a function of the azimuthally symmetric radial distance in inverse centimeters squared. The radius of gyration used for the plot was 1.35 nanometers. This is the standard shape of an azimuthally averaged Guinier plot.

Figure 3.5 demonstrates the resultant Guinier plot of a dilute, homogeneous solution of particles. To find the size of these particles, the plot is cropped and a linear fit is performed so as to solve eq. (3.36).

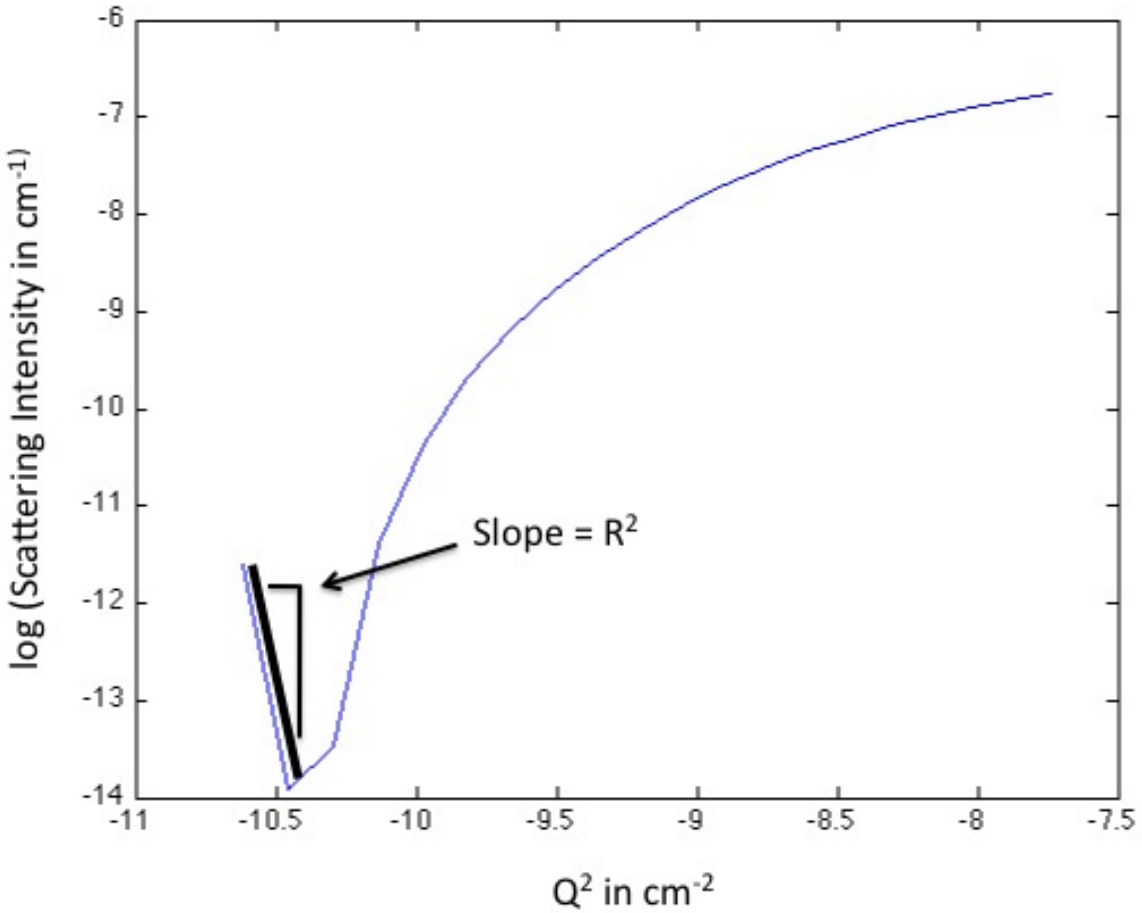


Figure 3.6: Theoretical plot of the natural log of the scattered intensity as a function of the azimuthally symmetric radial distance in inverse centimeters squared cropped to allow a linear fit to be performed to find the size of the particle. The most linear portion of the plot is found. Its corresponding slope is the relation R^2.

Error arises here due to several factors. A linear fit is used in Guinier analysis to fit data to a slope that is squared. If the solution is not completely homogeneous, there will be trouble fitting and finding a radius of gyrations because multiple radii would be contributing the slope in the linear fit. Radii of mixed populations could be more accurately calculated if more than one value, *R,* is incorporated into calculations. After analysis of actually data, further error discussion will follow in chapter 5 .

Chapter 4

Experimental Methods

Section 4.1: Overview

The following chapter will discuss the experimental setup and data collection involved in determining protein size using SAXS. From the generation of light to absorption of scattered x-rays off of a prepared sample, the experimental tools needed for scattering study will be outlined. Before the protein size can be determined, it must be prepared. An extensive overview of the step-by-step sample preparation will be discussed. Finally, data collection using the SAXS sampling setup at BioCAT and Igor Pro will be outlined. More detailed sampling techniques and data analysis procedures can be found in Appendix 4.

Section 4.2: Preliminary Laboratory Setup: X-Ray Generation

X-ray radiation is produced for experimental use at the Advanced Photon Source (APS) at Argonne National Laboratory (ANL). The APS utilizes synchrotron radiation to provide an x-ray source for data sampling.

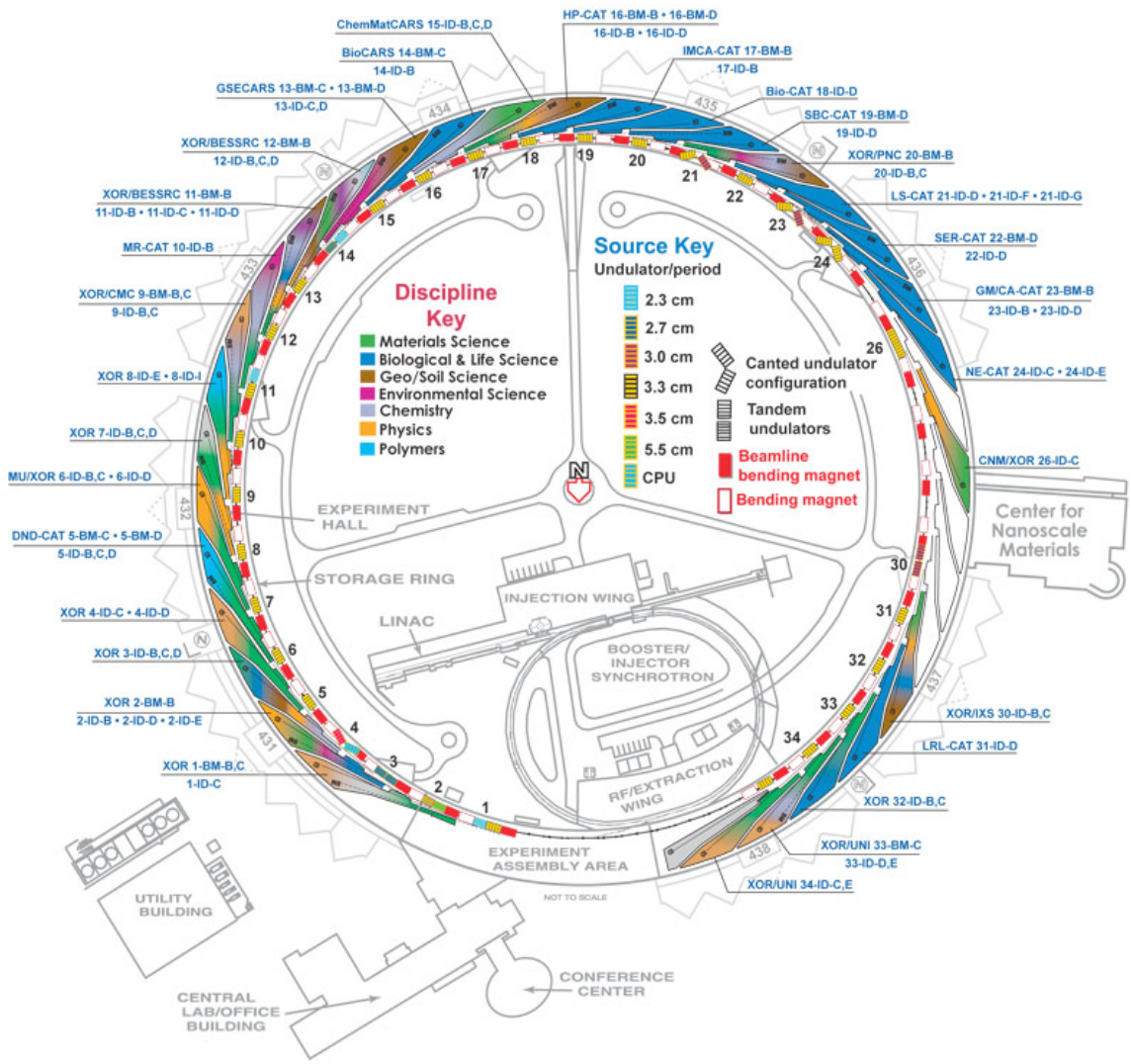


Figure 4.1: Diagram of linear accelerator, injector and storage ring. Compliments of http://www.aps.anl.gov/Beamlines/Beamlines_Map/index.html.

Electrons originate in the booster/injector labeled in Fig. 4.1. From there, they are accelerated in the linear accelerator and sent into the storage ring at 7 GeV. The APS consists of 40 different segments. 34 of the segments (or sectors) have their own undulators consisting of alternating permanent magnets, which force the electrons to undergo oscillations and emit radiation. This radiation is relativistically Doppler upshifted and collimated due to the high electron energy. At sector 18, where this experiment was performed, the light emitted from the storage ring is in the x-ray range with energies ranging from 3.5 to 39 keV. This experiment used x-rays in the range of 7 keV. These x-rays then travel via vacuum chamber through two Silicon monochromators to make the light nearly monoenergetic. The beam is then focused using toroidal mirrors. Using burn paper and in the safety of a lead lined hutch, x-rays can be focused on a sample to no less than approximately 60 square micrometers.

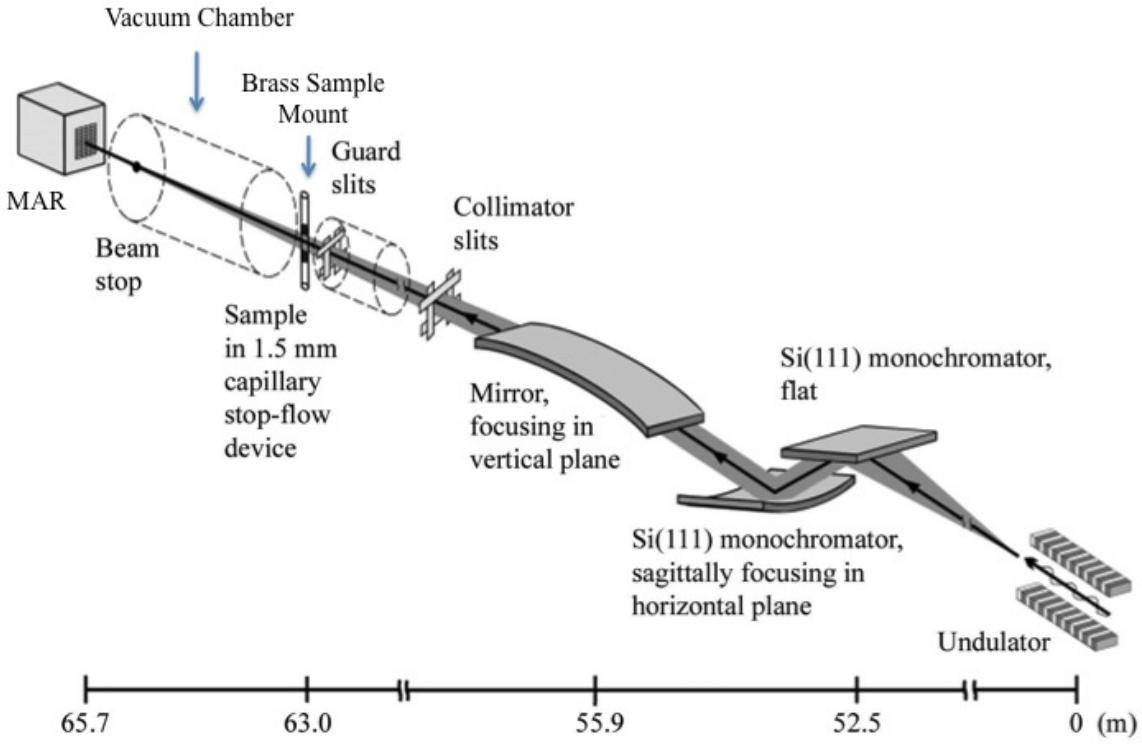


Figure 4.2: Experimental layout for x-ray beam collimating and focusing as well as data sampling (Fischetti 2004).

As the x-rays propagate from right to left in Fig. 4.2, the beam is focused through the guard slits onto the sample located in the brass sample mount. The brass sample mount is temperature controlled and monitored to vary sample conditions.

The beam is focused on a specially made brass, sampling mount and will exit its vacuum tube and enter the sample mount. Inside the mount, a 150 micron diameter capillary is held in place with hosing that suctions the sample in and pumps the sample out via a stop-flow, Microlab 600 pump and computer. The x-rays scatter off the sample held inside the capillary and reenter another vacuum chamber. At a length of one meter, this chamber provides an area for the x-rays to propagate undisrupted by particles in the air. This ensures that the only scattered intensity observed is that of the sample inside the capillary. At the end of this vacuum chamber is a MAR 100K x-ray CCD detector. This detector captures the scattered x-rays and sends them to a computer for image processing.

Section 4.3: Preliminary Laboratory Setup: Laboratory Modifications for Temperature Dependent Data Acquisition

Once the x-rays are collimated and focused onto the location of data collection, the brass sample chamber needs to be modified to account for temperature change in the system. Sector 18's standard SAXS data collection setup was utilized with a few modifications shown in Fig. 4.3 to allow for testing over a large temperature range. Sample collection is desired between

temperatures from -25 degrees Celsius to 50 degrees Celsius. A Thermo Neslab RTE-740 Digital Plus Refrigerated and Heating bath was attached to the sample mount using input and output hoses. The temperature bath was filled with a 60% mixture of automotive grade antifreeze to prevent freeze up of the solution at low temperatures. The remainder of the bath was filled with distilled water to prevent calcium build up throughout the bath, the hoses and the sample mount. A digital thermocouple or temperature probe was attached to the brass sample to better determine the sample temperature when exposed to x-rays. Data collected corresponds to this temperature reading. To prevent heat loss, the hoses are insulated with polystyrene. Likewise, the sample mount is encased in polystyrene to prevent thermal loss before reaching the protein sample. At temperatures at or below freezing, ice crystals can form from condensation in the air. Crystallization drastically reduces SAXS data quality by creating a strong ice diffraction pattern. To remove the possibility of ice formation, helium is blown into the sample mount at a rate that does not cause turbulent flow or jitter in the protein sample. With these modifications, the data collection can take place at temperatures at the ranges desired.

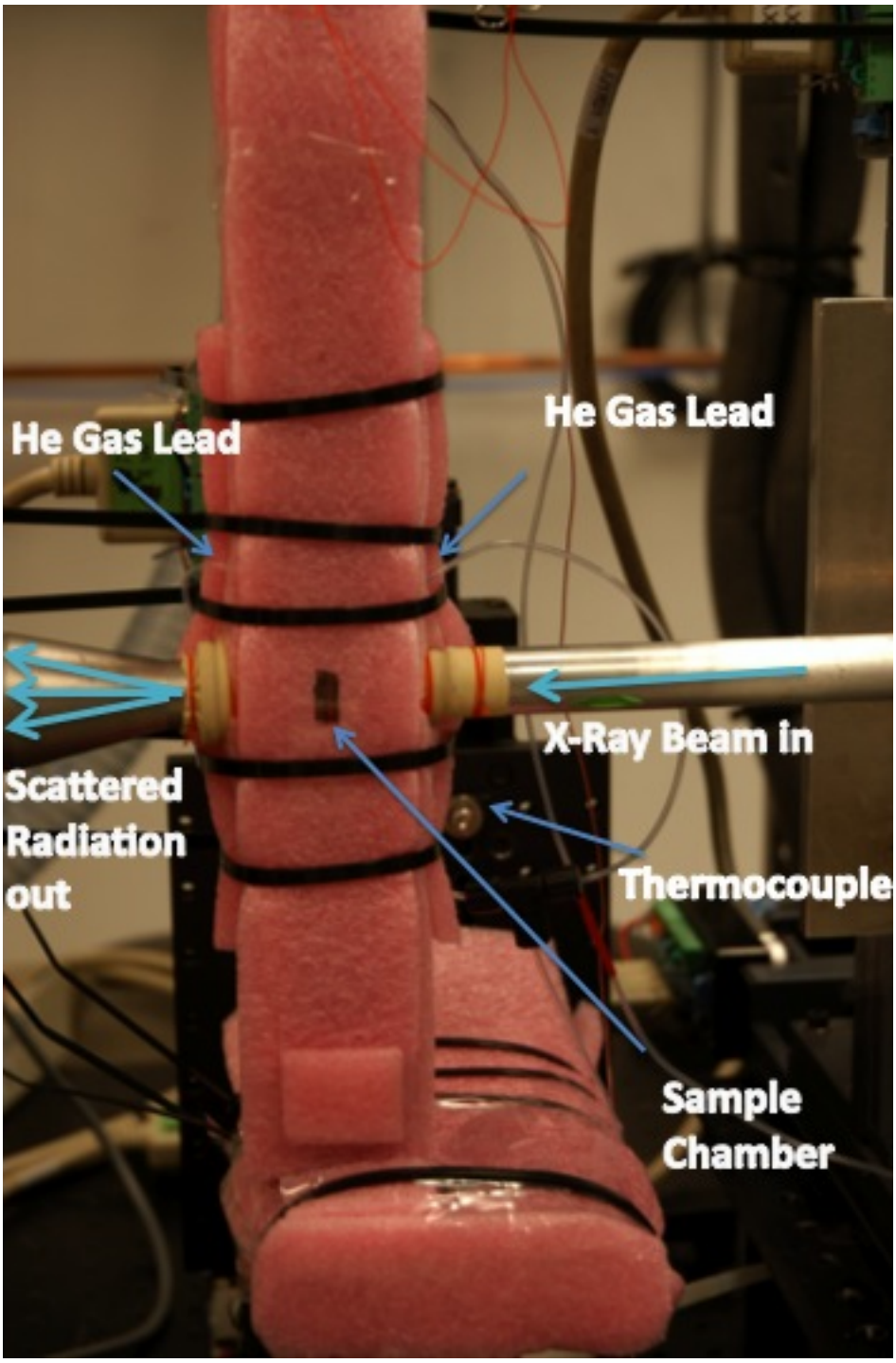


Figure 4.3: Modified sample mount thermally insulated and ready for data acquisition.

Insulation of the sample mount and hoses was not completely effective. Temperature variations between the reported temperature of the chilling bath and the thermocouple became very apparent at low temperature. Cooling the sample became increasingly difficult below -20 degrees Celsius, and below -25 degrees Celsius was impossible due to the maximum power output limit of the chiller.

With all temperature modifications in place, the experimental apparatus is ready for data acquisition. Before data can be collected, samples containing both solution and protein need to be prepared as described in the next section.

Section 4.4: Sample Selection: The Protein and Its Denaturant

Cytochrome-C was chosen as the protein for study in this experiment for a variety of reasons. Its size is well known and fits the size criteria desired for study. The protein's genomic sequence is documented for several variations of the molecule (Bilsel 2006). Further, the radius of gyration has been published for the native, denatured and two intermediate states (Akiyama 2001, Pollack 1999).Cytochrome-C folding has been widely studied using other forms of folding testing. Finally, Cytochrome-C is inexpensive and its central structure contains Tryptophan. Tryptophan responds to UV radiation by fluorescencing (Wu 2008). The fluorescencing of the protein can be indicative of folded position and can be used as a local probe to study the stage of denaturation of Cytochrome-C.

Protein State	Radius of Gyration	Method of Computation	Reference
Native State	[ ] Angstroms	Kratky Analysis (WAXS)	Pollack 1999
	12.8 – 13.5Angstroms	Guinier Analysis (SAXS)	Hsu 2007
Intermediate State I	20.5 Angstroms	Guinier Analysis (SAXS)	Akiyama 2001
Intermediate State II	18 Angstroms	Guinier Analysis (SAXS)	Akiyama 2001
Denatured State	30-32 Angstroms	Kratky Analysis (WAXS)	Pollack 1999
	29.5 – 29.7 Angstroms	Guinier Analysis (SAXS)	Hsu 2007

Table 4.4: Literature values of measured stages of denaturation of Cytochrome-C using various techniques. Both small angle x-ray scattering and wide angle x-ray scattering (WAXS) were methods of radius of gyration measurement.

Table 4.4 shows the size variation of Cytochrome-C from its native to its denatured state through its intermediates using varying techniques. Variations in radius of gyration size for each stage of denaturation are published depending on method sample preparation and data collection and measurement. All samples in Table 4.4 correspond to x-ray scattering measurements taken at room temperature and low pH of approximately 2.

Protein structure is dependent on the organism of origin. Equine Cytochrome-C, see Fig. 1.1 for structural characteristics, was used during data acquisition on both July 7th and December 18th of 2009. As of yet, specific equine genomic properties have not been published.

Cytochrome-C information for organisms close to that of the equine species have been used as needed and are dually noted when necessary.

Cytochrome-C is supplied in extracted, crystalline form. The protein must be redissolved into solution to prepare it for the experiment. A solution was chosen that would allow the protein to remain in solution at temperatures as low as -25 degrees Celsius and as high as 50 degrees Celsius. The solution would also need to mirror the natural occurring environment for Cytochrome-C. To achieve both of these requirements, the solution consists of a partial mixture of water and ethylene glycol with a low concentration of salt and commonly used Tris buffer. Ethylene glycol lowers water's freezing point, does not interact with the protein and will not affect the scattering of x-rays off of Cytochrome-C. Both Tris buffer and salt create a buffer that stabilizes the pH of the solution. The pH of the solution was chosen to be low to allow for comparison with published values. To achieve the appropriate pH, minimal volumes, approximately 50-200 µl per 200 ml solution, of 6 molar Sodium Hydroxide, NaOH, and 19 molar Hydrochloric acid, HCl were added to the solution. The exact formulation of the sample solution is detailed below.

This experiment relies on both temperature and denaturant concentration to unfold Cytochrome-C. Guanidine HCl, or GuHCl, is a commonly used protein denaturant. Guanidine HCl, dissolved from its crystalline form, provides an excellent tool for protein manipulation. The denaturant is reconstituted into a 6.6 molar solution. This high concentration allows for dilution to varying degrees.

Section 4.5: Sample Preparation: The Solution, Protein and Denaturant

To ensure consistent sampling, high standards in laboratory procedure are required. A clean environment and accurate instruments are needed to guarantee sample preparation precision and accuracy. Samples and solution were prepared utilizing the facilities at sector 18 of the Advanced Photon Source at Argonne National Laboratory. All materials were prepared in a wet laboratory setting where the rubber gloves and appropriate eyewear were used to shield from chemical contact and prevent contamination.

The solution needs to be tailored to permit testing above and below the freezing point of water. Ethylene glycol is not known to react with water, Cytochrome C or Guanidine HCl. Ethylene glycol lowers the freezing point of water to varying degrees depending on the percentage by volume added to water. A 45% ethylene glycol and 55% distilled water mixture was chosen because the solution freezing point is below -40 degrees Celsius.

100 ml 45/55 mix		
Volume	Chemical	MDL number
45 ml	Ethylene Glycol	MFCDOOOO2885
55 ml	Distilled H_2O	

Table 4.5: General solution recipe for all cleaning and sample solvents needed.

The solution recipe outlined in Table 4.5, referred to as 45/55 mix, provides the backbone for sample preparation and later, experiment cleaning. The 45/55 mix is prepared in bulk and will be used as the solvent for all other sample related solutions. This mixture is used in instrument cleaning as well. Care needs to be taken in the capillary cleaning of the brass sample mount at temperatures below freezing. The fragility of the capillary requires that no build up of ice occur in or around the capillary at any time. Freezing, in or around the capillary can cause damage to the laboratory setup. Replacement of a damaged capillary is a time consuming procedure that needs to be avoided when at all possible.

Once the basic solvent has been prepared, the sample solution must now be buffered. Buffering controls the pH of the solution, thus helping the solution maintain its characteristics at varying temperatures. For every hundred milliliters of 45/55 mix, 0.5844 grams of NaCl and 10 milliliters of Tris buffer were added. The solution was mixed using stirring rods on a stirrer/hot plate until the sodium chloride dissolved completely. After mixing, the buffered solution is ready for the protein and the denaturant.

110 ml Buffered solution at pH 3.52			
Mass or Volume	Chemical	Molarity	MDL number
100 ml	45/55 Ethylene Glycol/H_2O mix	N/A	See table 5.4
10 ml	Tris buffer	.1 M	NC95254 7 9
.5844 g	NaCl	.15 M	MFCD00003477

Table 4.6: The general buffered solution used to moderate equilibrium conditions of the aqueous sample.

Before data acquisition, the solution was mixed on hot plate/stirrer using stir bars until salt was completely absorbed into solution. Once crystalline salt was completely dissolved, the pH of the solution was measured and set appropriately. A pH of approximately 3.5 was chosen so that comparisons with published Cytochrome C data could easily be made. The radius of native state, both intermediate states I and II and the denatured state of state of Cytochrome C have been published at pH values of approximately 3.5 (Akiyama 2001, Pollack 1999). The prepared solution was tested using the Accumet Research AR15 pH meter provided by the BioCAT facilities. Before reading the pH of a given mixture, the meter must be calibrated. Specific pH calibration for the Accumet AR15 can be found in Appendix 5a. After properly calibrating the pH meter, the pH was adjusted using 19 Molar NaCl, 1 Molar NaCl and 5 Molar HCl to a final reading of pH 3.52. After the solution is prepared to a comparable composition, denaturant and protein need to be readied for sampling.

To place Guanidine HCl into solution, the appropriate amount of denaturant needs to be massed out to prepare a 6.6 molar solution. The denaturant is prepared in high concentration so that it may be mixed at will and diluted as needed. Guanidine HCl has a molar mass ratio of 95.53 grams per mole. To achieve a 50-milliliter solution at 6.6 molar concentration, 6.6 times the molar mass of GuHCl is calculated. This mass will yield a concentration of 6.6 molar GuHCl if added to fill 1 liter of solution. One liter of solution is a very large volume. A volume of 50 milliliters will prove to be a copious yet manageable volume to work with.

31.5 grams of crystalline Guanidine HCl were massed out and placed in a 50ml conical. The conical was then slowly filled to 50ml with a buffered 45/55 mix. It was occasionally necessary to close the conical and gently agitate the mixture. Guanidine HCl is very soluble in water but not in ethylene glycol. Because of this, 30 minutes was needed to get GuHCl into solution. Once all traces of crystalline Guanidine HCl appear to have dissolved, the conical was then completely filled to 50 ml and the meniscus was checked to ensure the appropriate volume was filled. The appropriate quantities of each chemical in highly concentrated Guanidine HCl solution as well as ingredient identification number, or MDL number, are contained in Table 4.7.

50 ml of 6.6 molar Guanidine HCl solution		
Mass or Volume	Chemical	MDL number
31.5 g	Guanidine HCl (GuHCl)	MFCD00013026
Fill to 50 ml in volume	45/55 Ethylene Glycol/H_2O mix	See table 5.5

Table 4.7: Guanidine HCl 6.6 Molar Concentrated Denaturant Solution used throughout the experiment.

The protein should also be prepared so it can be diluted and used as needed. Initial storage in a -20 ^oC freezer proved problematic. At low temperatures in low pH solutions, Cytochrome-C would unfold but not refold. To avoid taking invalid data, after the protein was put into solution, the samples were stored at room temperature. Proteins can be imaged using SAXS when in concentrations of at least two milligrams of protein in one milliliter of solution or higher. Choosing a higher concentration to prepare ensures that there is enough protein in solution for data analysis. Further, putting the protein into a solution that is ten times the concentration desired for testing will allow bulk solution to be made and diluted. Ideally, a 4-mg/ml concentration is desired for testing. Therefore, a 40-mg/ml solution should be prepared and diluted as needed. To prepare such a concentrated solution, 50 mg of crystalline Equine Cytochrome-C was massed out. The protein was then transferred to a 2 ml conical. Finally, 45/55-buffered mix was added to fill the 1 ml mark. The meniscus was inspected and the volume is noted. The appropriate quantities of each chemical in highly concentrated Cytochrome-C solution as well as ingredient identification number, or MDL number, are contained in Table 4.8.

40 mg/ml Protein Solution (Cytochrome-C Solution)		
Mass or Volume	Chemical	MDL number
50 mg	Equine Cytochrome-C	MFCD00012590
Fill to 1.25 ml in volume	45/55 Ethylene Glycol/H_2O mix	See Table 5.5

Table 4.8: Highly Concentrated Cytochrome-C in buffered solution.

Using both highly concentrated Guanidine HCl and Cytochrome-C in solution, samples at varying degrees of denaturation can be prepared. Taking SAXS on the 45/55-buffered mix and denaturant would reveal a scattering intensity curve due to the solution. Because of this, each protein solution must be prepared with a partner background solution containing the same

volume of denaturant but no protein. When data is taken, the background is run and its corresponding intensity curve is subtracted from the protein's intensity curve for a given run.

10 µl of concentrated protein solution were pipetted into a sampling conical. Depending on the denaturant concentration needed, the required Guanidine HCl is pipetted into the same conical. To dilute solutions to the desired concentration, 45/55-buffered mix is pipetted into the conical. For each data run, 100 µl of sample are needed. Five different concentrations of GuHCl, between 0 and 4 molar, were tested. Twenty samples of both the protein in solution and its corresponding background were prepared at a time. Preparing solution in bulk before taking data ensured coherency in data collection. Table 4.9 contains the needed ingredients and their associated concentrations for each desired protein sample denaturant concentration. Table 4.910 contains the needed ingredients and their associated concentrations for each desired background sample denaturant concentration.

4 mg/ml Protein in Denaturant					
	Volume of Solution for Given Denaturant Concentration				
Solution	0 M GuHCl	1 M GuHCl	2 M GuHCl	3 M GuHCl	4 M GuHCl
40 mg/ml Cytochrome C	10 µl	10 µl	10 µl	10 µl	10 µl
6.6 M GuHCl	0 µl	15 µl	30 µl	45 µl	60 µl
45/55 Ethylene Glycol/H_2O mix	90 µl	75 µl	60 µl	45 µl	30 µl

Table 4.9: Diluted Cytochrome-C in varying concentrations of Guanidine HCl

Background Denaturant Solution					
	Volume of Solution for Given Denaturant Concentration				
Solution	0 M GuHCl	1 M GuHCl	2 M GuHCl	3 M GuHCl	4 M GuHCl
6.6 M GuHCl	0 µl	15 µl	30 µl	45 µl	60 µl
45/55 Ethylene Glycol/H_2O mix	100 µl	85 µl	70 µl	55 µl	40 µl

Table 4.10: Blank solution of diluted Guanidine HCl

Once each solution was combined in a 2 ml conical, the resultant volume is agitated using a vortex to facilitate mixing. Several attempts at mixing using tabletop centrifuges resulted in complete separation of solution. Using a pipette to pump the volume in and out of the conical, thus mixing the solution, also failed. Pipetting the solution too fast may cause air bubbles. Air bubbles distort SAXS data. When an x-ray encountered an air bubble, the resultant scattering pattern stored the interaction as a spiked data point in the data reading. All data with air bubbles had to be removed because inconsistent spikes render the data inaccurate.

Section 4.6: Sample Loading and Experimental Setup Cleaning

Once both the background and the protein sample are prepared, data collection can begin. Data is first acquired for the background sample. The sample mount must be cleaned before every background sample and after every protein sample is run through the experiment. Data of the background is taken first as its scattering results are subtracted out of the protein's scattering data to acquire only the scattering pattern of the protein. Because of this, it is unnecessary to clean in between the background and protein sample exposure.

To clean the sample mount, a cleaning procedure has been designed to ensure the capillary and attached hosing have been completely evacuated of any residual sample. Five separate solutions were used to gradually clean the capillary. Three 50 ml conicals were filled with 45/55 mix, one conical was filled with isopropyl alcohol and one conical was filled with a 20% bleach solution. To initiate the cleaning process, one of the 45/55 mix conicals was placed into the brass sample mount. Approximately 3 milliliters of solution or two full revolutions on MicroLab 600 pump syringe were pumped through the sample mount using the manual control switch on the MicroLab 600 pump. Next, the conical of isopropyl alcohol was placed into the sample mount and approximately three milliliters or two full revolutions of the pump syringe were loaded. Using another 50 ml conical of 45/55 mix, a three-milliliter volume was loaded into the sample mount and pumped through the conical. The conical of 20% bleach solution was then loaded and run through capillary and sample mount using the same volume as with the first three cleaning solutions. Finally, the last conical of 45/55 mix was loaded into the sample mount and run through the system. To prepare for data sampling, the entire sample mount, including pump and hoses, needs to be evacuated. After the last conical was loaded and the appropriate volume had been suctioned into the system, the conical was removed but the pump was left on until the system was evacuated.

With a properly cleaned sample mount, the background can be loaded and data can be taken. To load the background sample into the sample mount, the 100 µl background sample contained in a 2 ml conical was inserted into the sample mount with the capillary hosing touching the bottom of the inside of the conical. After selecting the GL-Protein program on the MicroLab 600 computer, the sample loading process could begin. Following the confirmation sequence through the program and then selecting to run the pump sequence prepared the sample to be exposed in the sample mount capillary. Further triggering is needed to initiate the pump sequence of the sample but this cannot be done until the x-ray beam is exposed to the sample and the MAR camera is triggered to take scattering data.

Section 4.7: Data Acquisition

After the sample is loaded into a clean sample mount and the computer has been triggered to move the sample in and out of the x-ray beam during exposure, the acquisition process can begin. Before exiting the experiment room or hutch, a camera aligned to view the MicroLab 600 controller was checked to ensure the program run could be monitored from a safe,

exterior vantage point. The hutch was then be searched following the safety protocol established for sector 18. Upon the complete evacuation of the hutch, the lead doors were closed and the x-ray shutter was opened.

For each background and protein data collection, a file name was chosen. The naming corresponded to the molarity of the denaturant in the sample and the temperature at which the samples were taken. Two computers were used to synchronize data collection and saved information under these corresponding names. On the processing computer, the x-ray exposure into the sample was monitored and recorded. Deviations in the sample, bubbles, and determination if the shutter was open could be seen through this monitoring program. After the sample file names for the run had been established, the monitoring program on the processing program was triggered by pressing "start" and a baseline was taken. This computer was now ready for data collection. The second, image acquisition, computer directly triggers the MAR detector and is set to acquire the scattered data. For each data run, 15 sample shots of x-ray scattering are taken. Each background and protein sample run will have a total of 30 images associated with the file, 15 for each sample. After coordinating file names, image number was set to 1 if background data is being run or 16 if protein data is being run. Images 1-15 pertain to background scattering patterns and images 16-30 pertain to protein scattering patterns. Setting the correct file sequence is the last step before actual data collection occurs.

To sample data, the MicroLab 600 computer must be triggered once more to start pump flow. To do this, a triggering gun, located exterior to the hutch was used to initiate the sequence. This triggering was coordinated in unison with the initiating of data collection from the MAR camera on the image acquisition computer. Data is then collected in 15 images for the sample loaded in the mount. The pump was timed to perform a constant flow pump of solution over the duration of image sampling and then deposit the exposed sample back into the 2 ml conical. Once data is finished collecting, the beam shutter could be closed and the sample could be removed and saved for potential later testing. Depending on the substance removed, the mount was cleaned and a new sample was loaded. After data for both the background and protein was collected, it could be analyzed to determine if another sample of the same constitution need be rerun. In the case that all data necessary had been collected at a certain temperature, the temperature of the system could be readjusted and sampling at a new set temperature could occur.

Section 4.8: Temperature Variation

The system's temperature could be manipulated at will to test samples at a variety of temperatures. To manipulate the temperature of the system, the Thermo Neslab RTE-740 Digital Plus Refrigerated and Heating bath was manually set to the desired temperature and a thermocouple was used to monitor when the system reached equilibrium. To change the temperature on the Neslab, the "Enter" button was pressed and the digital scale would start to flash. This flashing indicated that a new temperature could be entered. Using the "up" and "down" arrows on the display, the desired temperature could be chosen to an accuracy of a tenth of a degree. Pressing "Enter" again would set the temperature for the cooling device and thus the system.

Section 4.9: Initial Data Processing

After data was acquired, it is necessary to immediately perform an analysis of the scattering patterns obtained. Data analysis was then repeated after leaving the facilities. Before data could be reduced, data must first be taken consistently and in such a manner that the integrity of the sample of both the background and protein solutions is not compromised. The samples taken satisfied values well within accepted error as will be examined next.

Chapter 5

Experimental Data Analysis

Section 5.1: Analyzing Data

The following chapter will discuss the experimental data analysis of the scattering patterns collected during the December 16th-21st run at Argonne National Laboratory's Advanced Photon Source. Data was analyzed onsite as acquisition was completed and again after all data had been collected and the experiment had been completed. Data reduction from its scattering pattern form to the extraction of purely protein scattering information will be shown and discussed. Step by step, the path to finding the radius of gyration of the protein using Guinier analysis will be shown for the native, partially unfolded and denatured state of Cytochrome C will be shown. While most steps will be covered a more detailed analysis techniques and procedures utilizing the Igor Pro program and associated macros can be found in the Appendix.

Section 5.2: First Steps to Data Analysis

Data analysis starts as soon as a complete set of 15 background and 15 protein scattering samples have been acquired. The same data analysis procedure took place immediately after the data was acquired and again after the entire volume of data had been collected. To start analysis, the data from the processing and imaging computers were transferred to the same file. Using Igor Pro, the data from both independent filing sources was reduced to 30 data files corresponding to each scattering shot taken using the MAR detector. The data was then plotted using a Macro explicitly created by Liang Guo, Sector 18, APS, ANL. This Macro can be found on the BioCAT, Sector 18, website and is free to the public (BioCAT). Using this program installation, all thirty files could be graphed in two dimensions.

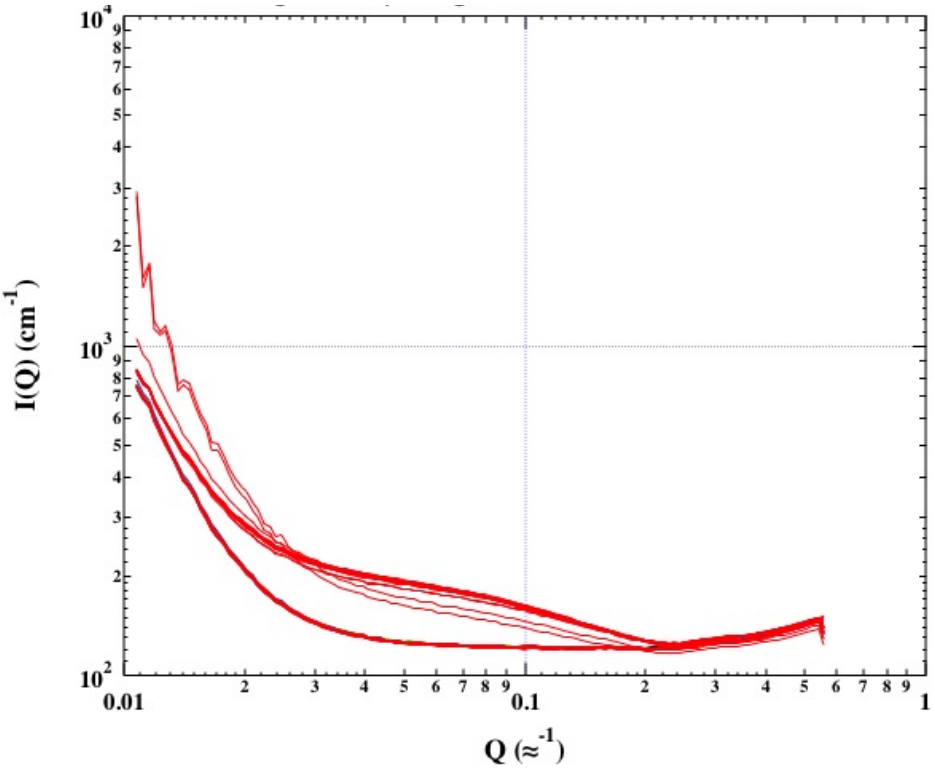


Figure 5.1: Raw data of Cytochrome-C in 2M Guanidine HCl and background 2M Guanidine HCl graphed in Igor Pro. Q is the log of the radial distance in centimeters from the x-ray beam. This Q corresponds to the Q defined in the initial theory section of this work.

As each scattering pattern is detected, a two dimensional, radially symmetric image is saved associated with the reading. Igor pro then takes the image and azimuthally averages the x-ray intensity at each radial distance. In this case, 120 distances along the radius of the scattering pattern were averaged. Figure 5.1 demonstrates what the raw data looks like before it is refined and the radius of the scattering object is calculated from it. 30 samples total appear in Fig. 5.1 corresponding to the 30 samples acquired during sampling. Two basic groupings appear in Fig. 5.1. The lower grouping of azimuthally averaged data points shows the background data sampled. The upper grouping of azimuthally averaged data points shows protein data sampled. Refining the sample by using Igor Pro to remove outlying data samples can lead to a better measurement of the radius of gyration of the protein.

Data samples that do not form a cohesively bunched sample can be removed to better measure the radius of gyration. Igor Pro allows for the deletion of samples that are of obvious error. Graphed samples with intensities of above 10^3 can be considered outlying. This intensity increase in x-rays can be attributed to poorly distributed sample or inhomogeneity in the sample as it is pumped through the capillary. Because samples are prepared before use, separation may occur in the solution. In several instances, it was noted that in solutions with Guanidine HCl, the Guanidine, a murky, viscous solution, separated from the 45/55, buffered mix, a clear, thinner solution.

Vortexing, or agitation by mixing of the sample before loading into the sample mount aiding in reinstituting homogeneity into the solution but it could not be guaranteed that the sample stayed well mixed once loaded into the hosing and capillary. For large proteins, vortexing may lead to destruction of the molecule. Because Cytochrome-C is a smaller protein, it is

possible to vortex or even centrifuge the protein for long periods (hours) without adverse affects to the protein itself. Vortexing before loading the sample should be no longer then 10 to 20 seconds though. This short time scale allowed any air bubble to exit the solution. Long-term vortexing showed visible air bubble formation for longer periods of time. Time periods of one minute or more created large air bubbles in the sample and on occasion the solute visibly separated from the solvent.

Even after being vortexed, the sample may not be completely mixed. Pockets of higher populations of protein may show up in the initial data analysis. These concentrated protein pockets or even Guanidine HCl pockets may be removed from the data during analysis. The standard practice is to eliminate the correlated data of abnormally high or low intensity as well as any data associated with no x-ray exposure or that of x-ray diffraction off of an air bubble. Air bubbles are evident through the observation of spikes in a usually smooth curve.

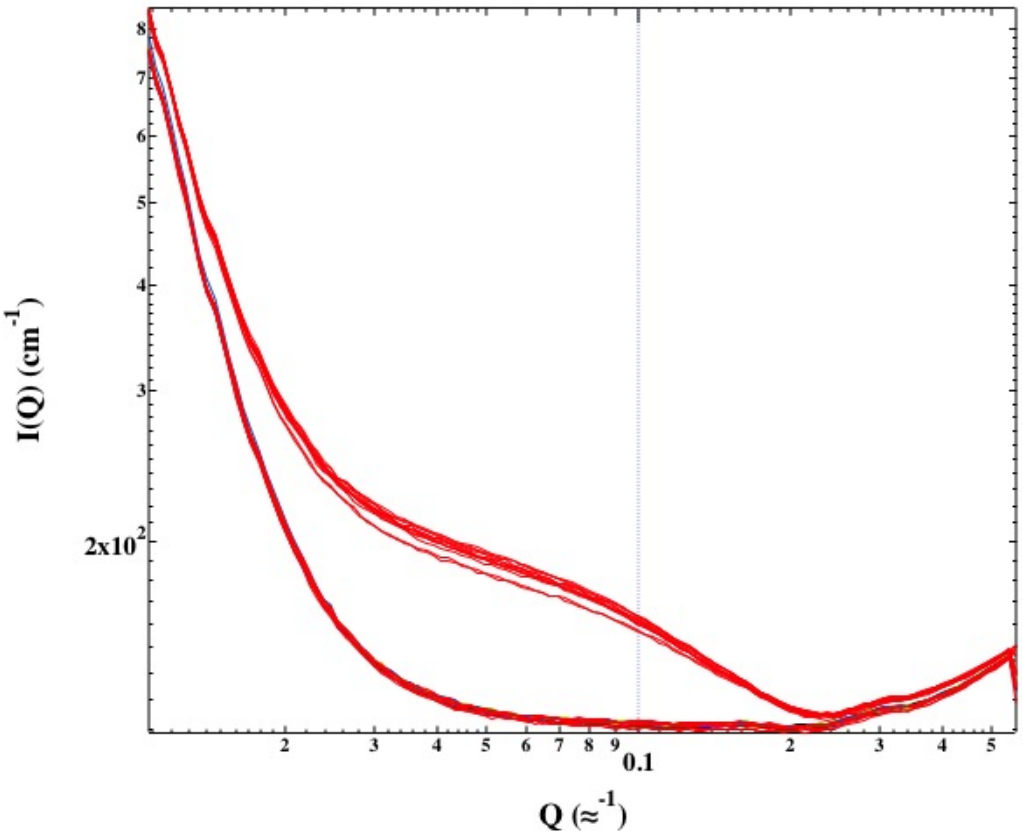


Figure 5.2: Cytochrome C in 2M Guanidine HCl and background 2M Guanidine HCl graphed in Igor Pro with outlying data removed. Q is the log of the radial distance in centimeters from the x-ray beam.

Once the ambiguities in the sample run have been removed, the log intensity verse log distance plot more clearly shows the distinctive difference between the protein solution scattering pattern and the background solution scattering pattern. Averaging the remaining data for the background provides a mean scattering intensity verse radius relationship for which to compare to the protein sample. Averaging the remaining data for the protein sample provides a mean scattering intensity verse radius relationship for which to subtract out the background from.

Subtracting out the background solution from the protein solution will provide a scattering plot of only Cytochrome-C. Igor Pro provided the means for averaging both the background and protein and data subtraction. These steps are detailed in Appendix 5. After subtracting the averaged background data from the averaged protein sample, the general shape of Cytochrome-C becomes quite obvious.

Section 5.3: Determining Protein Phase Structure

The direct observation of Cytochrome-C's scattering pattern can be seen in the plotting of background subtracted protein data in the same log intensity verse log distance plotting as used for figures 5.1 and 5.2.

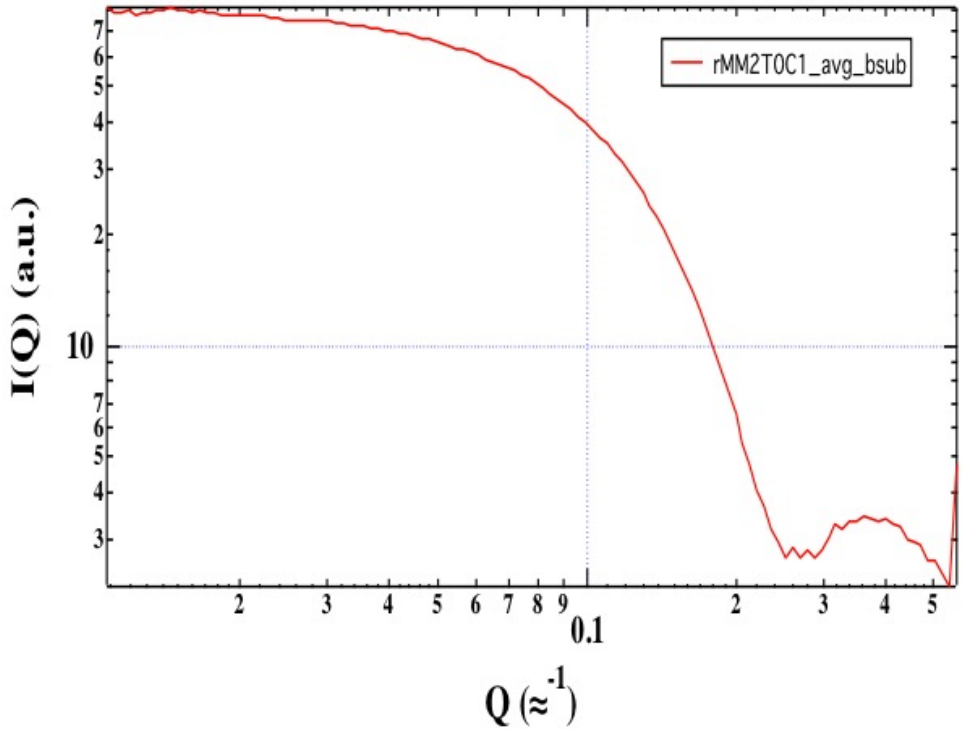


Figure 5.3: The log intensity verse log distance plot of Cytochrome- C's scattering pattern in 2 molar Guanidine HCl generated from subtracting the averaged background scattering data from the averaged protein scattering data.

Figure 5.3 shows the cross-sectional scattering pattern for Cytochrome-C alone. This plot is similar to that of the Fourier transform of a sphere seen in Chapter 3, Fig. 3.3. The width and curvature of the plot are directly dependent on the radius of gyration. Figure 5.3 demonstrates Cytochrome-C in a partially denatured state. The graph demonstrates the characteristics of that of a spherical Fourier transform with some deviation. As will be demonstrated later, for an approximately spherical object, the plot line for values of Q less than 5 is generally smooth. This smoothness occurs as the object moves further and further away from a spherical description or,

in this case, the protein unfolds. A partially unfolded state can be further verified in the measurement of radius of gyration.

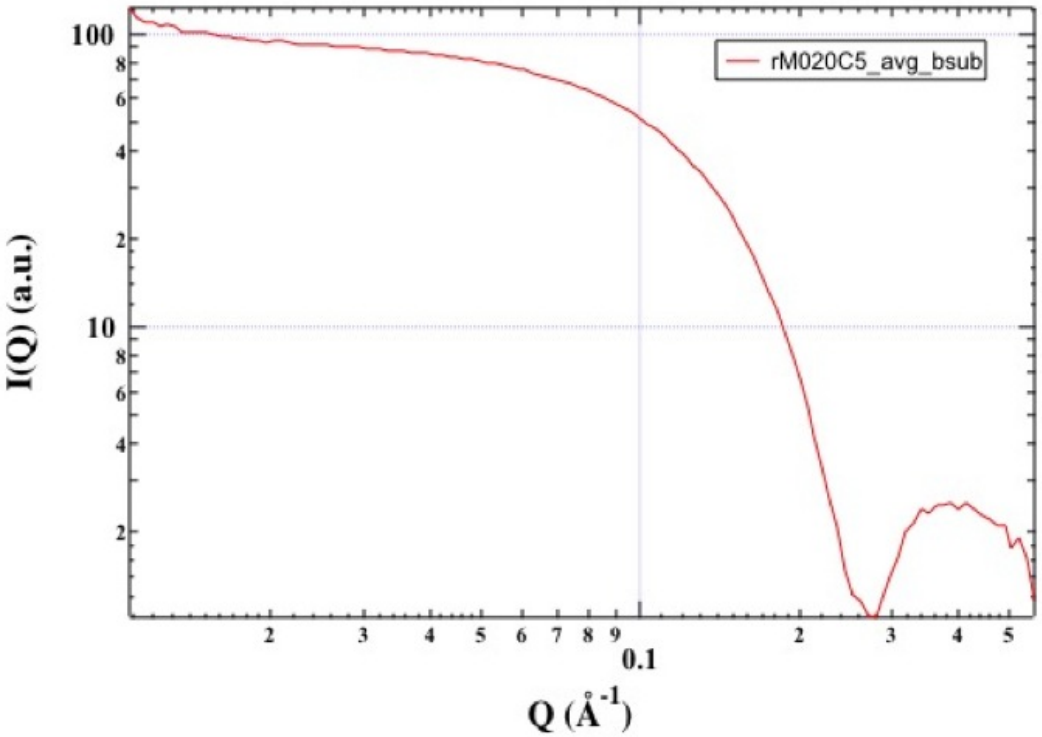


Figure 5.4: The log intensity verse log distance plot of Cytochrome C's scattering pattern in 0 molar Guanidine HCl generated from subtracting the averaged background scattering data from the averaged protein scattering data

.

Dependent on the size of the protein, the subtracted plot will have a broader or narrower curvature. At varying temperatures and concentrations of denaturant, these plots should reveal curvature variations dependent on radius of gyration of the folded or unfolded protein. The Protein scattering curve in 0 molar denaturant in Fig. 5.4 expresses characteristics of a spherical object much like that of the Fourier transform prediction. As the protein unfolds, it becomes less spherical and its scattering plot loses the characteristics of a spherical object. As Cytochrome-C denatures, it the radius of gyration should increase and the protein elongates and loses the spheroid structure.

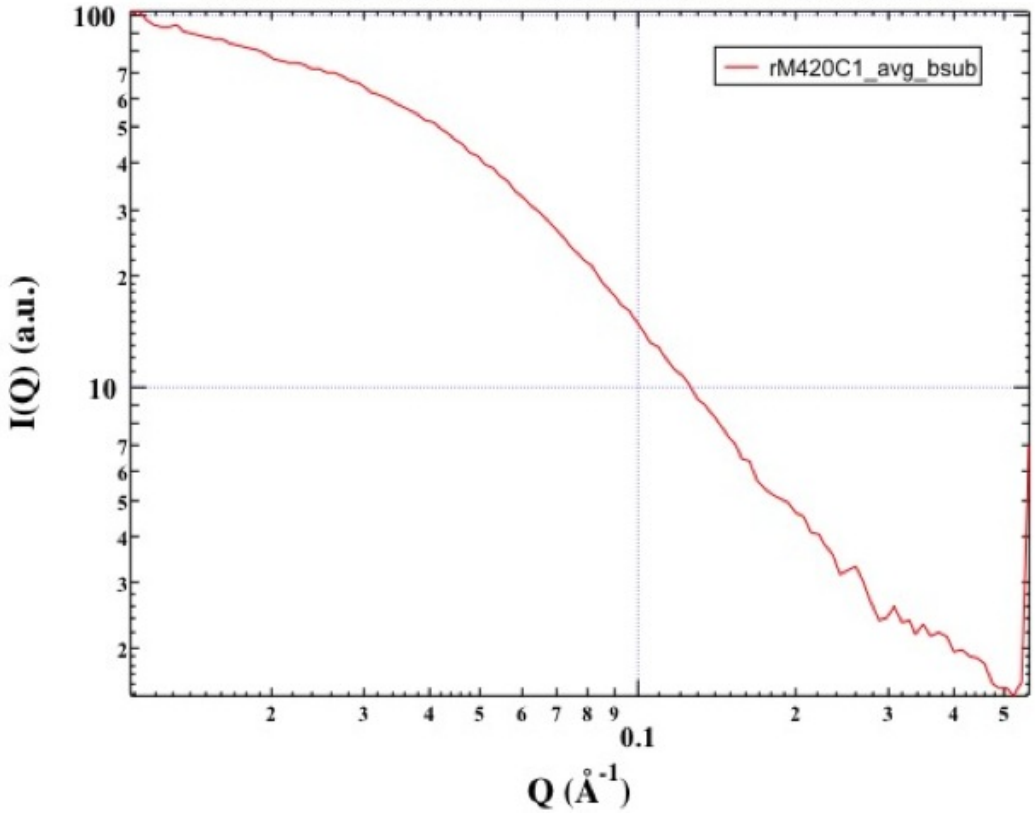


Figure 5.5: The log intensity verse log distance plot of Cytochrome-C's scattering pattern in 4 molar Guanidine HCl generated from subtracting the averaged background scattering data from the averaged protein scattering data.

The scattering pattern for Cytochrome C in 4 molar denaturant in Fig. 5.5 reveals a curve that has lost the characteristics of the Fourier transform of a sphere. The unfolded protein no longer adheres to a spherically compact modeling. Regardless, using Guinier analysis, the radius of gyration of the macromolecule can be found.

Whether the protein is compactly folded in a native conformation or unfolded in a denatured state, Guinier analysis permits the accurate measurement of the radius of gyration of the molecule. While the denatured molecule may or may not be a compact spheroid it will display a characteristic, larger, radius of gyration in comparison to the native state. This occurs because the unfolding process eliminates the alpha helices and beta sheets of the molecule allowing for either a more linear structure or a random coil where complete molecular linearity has not been attained. A linear, unfolded protein may still be treated as a sphere because through SAXS, its major axis of rotation or radius of gyration can still be determined.

Section 5.4: Determining Protein Size

A folded, partially unfolded, or completely unfolded state can be further verified in the measurement of radius of gyration. Using the scattering theory derived in chapter 3, Igor Pro

uses Guinier Analysis to find the radius of gyration, R_g, from the scattering pattern taken by the MAR CCD.

As outlined in chapter 3, the radius of gyration is solved for using the portion of the scattering pattern nearest to the beam block and linearly fitting it with the Igor Pro Guinier plot command. Using the averaged data a Guinier plot is formed by taking the log of the intensity and plotting it against the square of the intensity's location. The graph is linear nearest to the beam source. A perpendicular deviation arises in most Guinier plots from error in data acquisition. Removal of this perpendicular error is intrinsic to the scattering pattern. Incorporating a larger beam block into the experimental setup may remove this error but could also hinder the measurement of the protein's size. As outlined in detail in Appendix 5, to actually measure the radius of gyration of the sample protein, markers are used to select the linear most section nearest to this perpendicular deviation on the Guinier plot. Guinier analysis is then performed using Igor Pro and a radius of gyration fit, R_g, is determined as well as its error and the maximum angle of the scattered intensity, [ ]. The error given is the error in linear fit to the points selected.

Figure 5.6 demonstrates a Guinier analysis on a Guinier plot of 0 Molar Guanidine HCl at 20 degrees Celsius. The measured radius of gyration for the protein only sample is approximately that of known recorded values of Cytochrome-C. In the attempt of determining either of the two intermediate states known to occur in the denaturation process, error in the radius of gyration of more than 1 Angstrom was considered unacceptable during data collection. This maximum error limit would allow for the determination of whether the sample was in the first or second intermediate state because their size is separated by approximately 1.5 Angstroms. All data collected and fit using Guinier Analysis have errors of equal to or less than 1 Angstrom. Most of the data collected had associated error in the linear fit of less than 1% from the measured radius of gyration. Error in the radius measured must come from somewhere else. Sample preparation improved over the course of the experiment but deficiencies in sample inventory as well as limitations in sampling time led to an inability utilize more improved techniques leading to fewer than desired high quality samples.

Figure 5.6 also demonstrates the maximum angle measured in small angle x-ray scattering. Recall that the scattered intensity is dependent on [ ] where [ ] is the vector that describes the angle of scattered intensity. Ignoring the constant, 2π, the value, [ ], has a desired set limit

[ ] (5.1)

. While the [ ] value exceeds the upper limit of eq. (5.1), it is very close to the maximum desired value and was accepted because of its well calculated radius of gyration and low error fit.

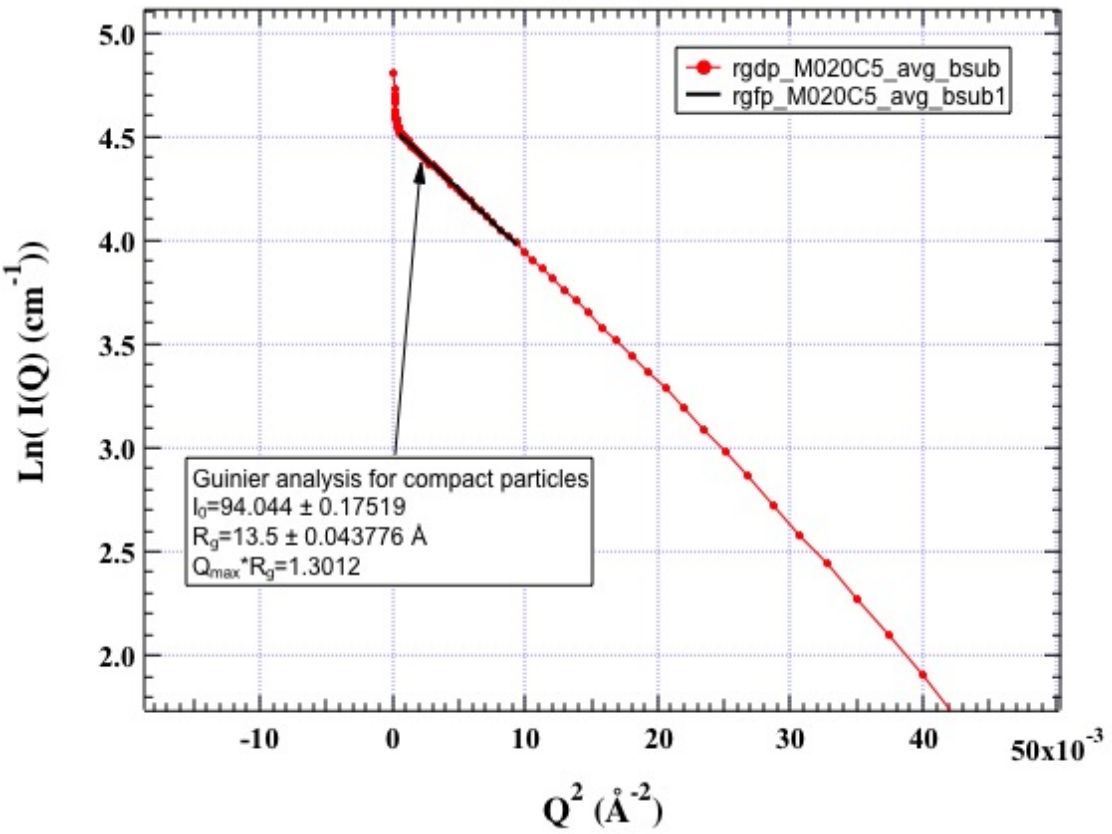


Figure 5.6: The log intensity verse radial distance squared plot, or Guinier plot, of Cytochrome-C's scattering pattern in 0 molar Guanidine HCl generated using the associated scattering data. The plot has been fit using Guinier analysis via the Igor Pro macros created by Liang Guo.

Figure 5.7 demonstrates Guinier Analysis of the Guinier plot of a Cytochrome-C sample in 2 Molar Guanidine HCl. In a pH of 3.52 and low concentration denaturant, the radius of gyration found closely resembles that of the second intermediate conformation radius of gyration. The measured radius of gyration is within the stated error limits for both R_g and [ ] showing that error in radius of gyration does not come from fit. It is possible that the sample is not homogeneous and is a mixture of many different conformational states. The radius of gyration reported does not take into account multiple conformations. The analysis could be readdressed to take several radii into account. However, here we limit our discussion to measuring the mean R_g .

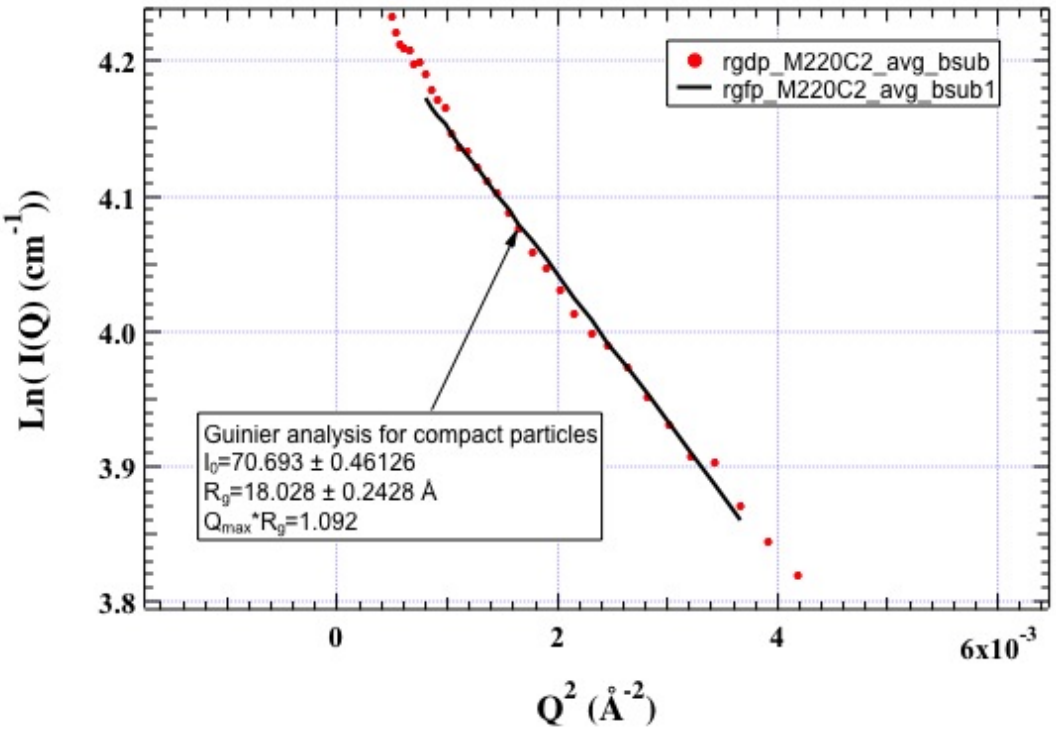


Figure 5.7: The log intensity verse radial distance squared plot, or Guinier plot, of Cytochrome-C's scattering pattern in 2 molar Guanidine HCl generated using the associated scattering data. The plot has been fit using Guinier analysis via the Igor Pro macros created by Liang Guo.

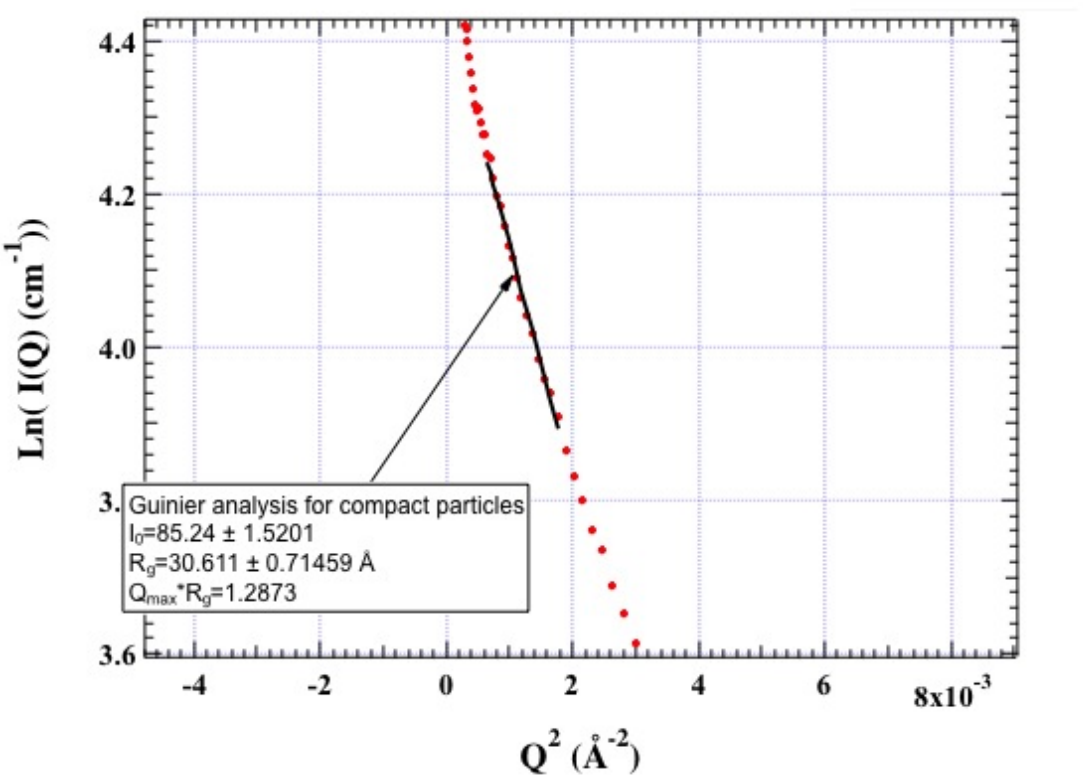


Figure 5.8: The log intensity verse radial distance squared plot, or Guinier plot, of Cytochrome-C's scattering pattern in 4 molar Guanidine HCl generated using the associated scattering data. The plot has been fit using Guinier analysis via the Igor Pro macros created by Liang Guo.

Figure 5.8 demonstrates that even close to the accepted error limit, a radius of gyration close to accepted value may be measured. The error in linear fit is minimal in comparison to the actual measured value. To ensure accurate measurement, the data collected was reanalyzed after experimentation was completed. Here, a radius of gyration of 30.6 ± 0.7 Angstroms is measured. This agrees with the accepted radius of gyration values reported in Table 4.4 of 29.5 – 32 Angstroms for a given technique. Further the radius of gyration measurement agrees with the SAXS determined mean R_g from Table 4.4 of 29.7 Angstroms when the error in fit is taken into consideration.

Error fitting in Guinier analysis also takes into account x-ray counting statistics. These counting statistics do not contribute significant error due to the high brightness of the x-ray source. The major source of error in these measurements is sample repeatability, which leads to uncertainties of approximately one Angstrom in size determination.

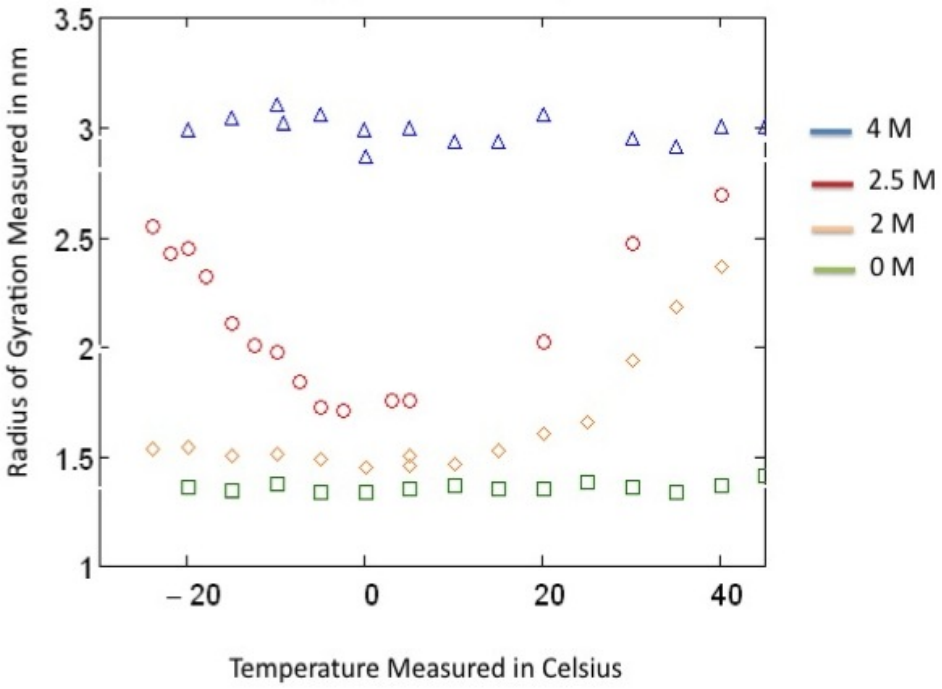


Figure 5.9: Radius of Gyration verse temperature plot of actual data taken during the experiment and reanalyzed. Error bars are not visible due to the error permitted being much smaller than the scale of measurement.

Appendix 5 contains the radius of gyration measurements as well as their errors during and after the data was taken. The data are further separated by molarity and temperature. Comparison to the Folding Free Energy developed earlier will reveal how competent data collection and analysis was and where further error lies. Error bars in figure 5.9 are non-existent for several reasons. Measuring the radius of gyration corresponding to each scattering pattern acquired and using the Guinier fit to measure error would result in error bars much smaller than the markers used to denote the average measurement recorded in figure 5.9. Error in data acquired at the Advanced Photon Source must come from other sources.

Section 5.5: Implementing Thermodynamic Derivation to Model Cytochrome-C

This section will implement the theoretical model derived in chapter 2 to model Cytochrome-C. Several set backs and assumptions will be discussed in the production of a denaturation curve model and the variables acquired to fully utilize the derivation.

Section 5.5.1: Basic Population Modeling of a Mixed Sample Population

The free folding energy model must some how be related to the radius of gyration of the protein in solution. To do this, the population in solution has to be considered. Using a two state model requires that a population be modeled such that at radius of gyrations that do not

correspond to that of a folded or unfolded state be modeled as a mixed population in the free folding interpretation. The two state model could be expanded to fit the known intermediates of Cytochrome-C but, as will be shown, within error, this is not necessary for understanding folding and cold denaturation of the protein.

A two-state model can be formed by making the assumption that out of 100% of the population, a protein can either have a folded structure or a denatured structure

(5.2)

.Where, p_n, is the population of proteins in the native state and p_d is the population of proteins in the denatured state. As the temperature is changed, dependent on denaturant concentration, the population of native to denatured proteins also changed. Temperature can then be incorporated to accurately describe the population change as a function of temperature for each denaturant concentration.

One way to restate unity is to separate it into a fractional sum

(5.3)

.The denatured and native proteins can be associated with each portion of this sum. The values of the proteins can be rewritten by incorporating the Boltzmann factor. The Boltzmann factor is a weighting factor that allows a variable to be written in terms of its probability of being in one state of another. Incorporating this factor into the population description, the components of the population can be rewritten as

(5.4)

(5.5)

.To incorporate temperature dependence, the variable *x*, can be related as the folding free energy divided by Boltzmann's Constant per mole, R, and temperature. The Boltzmann factor is used in thermodynamic systems. Taking into account the thermodynamic derivation in Chapter 2 for the Folding Free Energy will satisfy the use of the Boltzmann factor. This leads to the direct substitution of the Folding Free Energy divided by energy per mole, for the variable x

(5.6)

.The protein populations as a function of temperature are then

(5.7)

(5.8)

.The protein populations can now be used to relate the folding free energy to the radius of gyration of the protein. The radius of gyration measured is the sum of the denatured population multiplied by the radius of the denatured protein and the native population multiplied by the radius of the native protein

(5.9)

. For each concentration of denaturant, a curve can be derived to theoretically determine the radius of gyration at a given temperature.

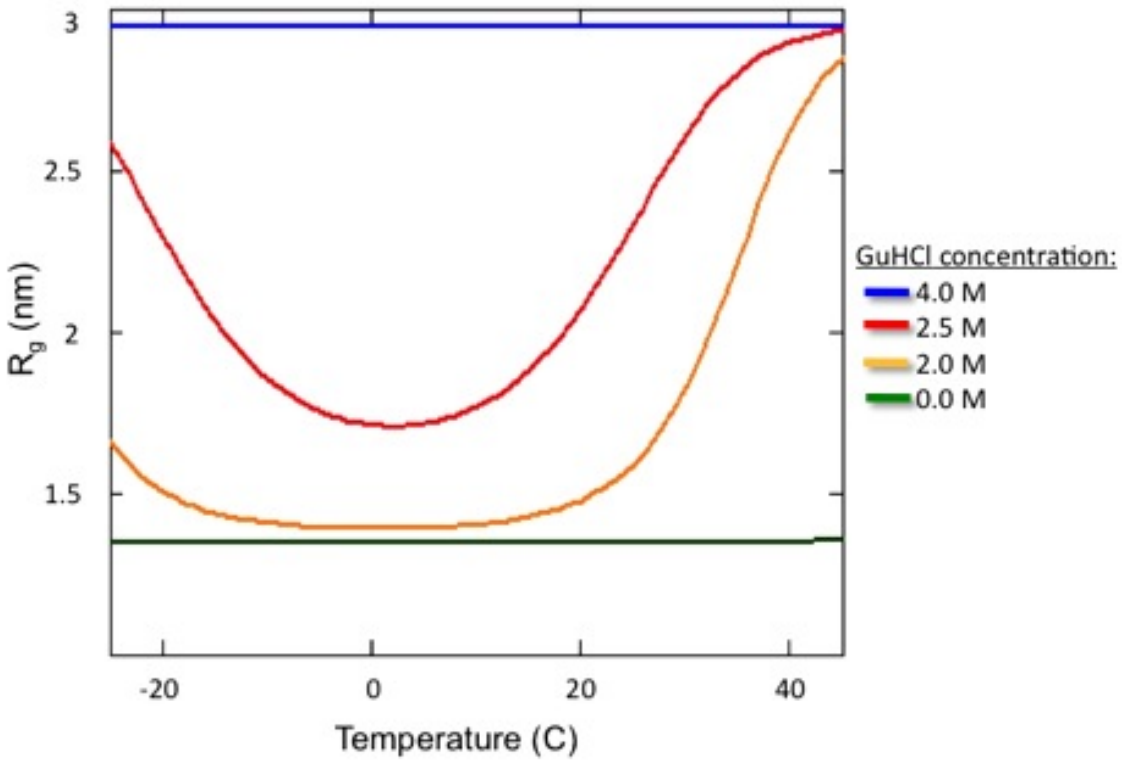


Figure 5.10: Theoretical Calculation of Radius of Gyration verse Temperature in Celsius for Cytochrome-C in Guanidine HCl concentrations of 0 molar, 2 molar, 2.5 molar and 4 molar.

For each molarity tested at the BioCAT Facilities at the APS, the corresponding radius of gyration at an associated temperature curve can be calculated by incorporating the free energy of

the protein at a given temperature. The ability to relate the folding free energy to the radius of gyration by implementing the population model makes prediction of a protein's size knowing its given conditions. Comparing the theoretical radii of gyration to the actual measured radii of gyration will demonstrate the authenticity of the two state model and all assumptions.

To construct fig. 5.10 several variables were needed

Variation	Definition	Value	Reference
g_o	Intrinsic Value of average packing energy of amino acid	-1200 cal/mol for	Ghosh 2009 Ideal Thermal Protein
□	Tested denaturant weighting factor	25 cal/mol-M	Ghosh 2009 for Guanidine HCl
c	Concentration of denaturant in solution	1,2,2.5, and 4 molar	Experiment Guanidine HCl
N	Number of amino acids in protein	104	(vender) Sigma Aldrich
z	Number of rotational Isomers in a given backbone	7.54	Ghosh 2009 Ideal Thermal Protein
□	Specific heat for protein	5.3 kJ K^{-1} mol^{-1}	Robertson 1997
T_h	Temperature at which enthalpy of the system vanishes	373 K	Ghosh 2009
T_s	Temperature at which entropy of the system vanishes	385 K	Ghosh 2009
R_n	Radius of gyration for the folded state	1.39 Angstroms	Pollack 1999
R_d	Radius of gyration for the denatured state	3.0 Angstroms	Pollack 1999
ε	Dielectric constant of sample solution substituted with the dielectric constant for water, see Appendix 5 for complete derivation of temperature dependent fit.	□	"Table 1", Dielectric Constant
pH(T)	pH of 45/55 buffered mix	□	Temperature dependent measurements performed in the lab

Table 5.11: Variables needed to construct Fig. 5.10 using the complete folding free energy equation derived in chapter 2 and the population model derived in chapter 5.

Charged Group for Cytochrome-C	pK_{exp}	Reference
His18	2.40	Wisz 2003
His26	2.90	Wisz 2003
His33	6.40	Wisz 2003
Tyr67	11.00	Wisz 2003
Lys79	9.00	Wisz 2003

Table 5.12: pK_i values needed to calculate localized protein charge, Q, for Cytochrome-C used to construct Fig. 5.10 using the complete folding free energy equation derived in chapter 2 and the population model derived in chapter 5.

Tables 5.11 and 5.12 record values needed to implement eq. (2.117) and calculate the total folding free energy for Cytochrome-C. At each temperature, the associated, calculated total folding free energy was used in the population model to form fig. 5.10. At each of four molarities, 0 molar, 2 molar, 2.5 molar and 4 molar denaturant, the radius of gyration of Cytochrome-C can now be predicted. Contrasting it with actual data taken during experimentation will demonstrate the validity of the theoretical model as well as the quality of the experimental data.

Section 5.6: Comparison and Contrasting of Data to Theoretical Modeling

Data collected can now be compared to the theoretical model derived to determine the precision and accuracy of the experiment. To do this, figures 5.9 and 5.10 can be combined into one plot.

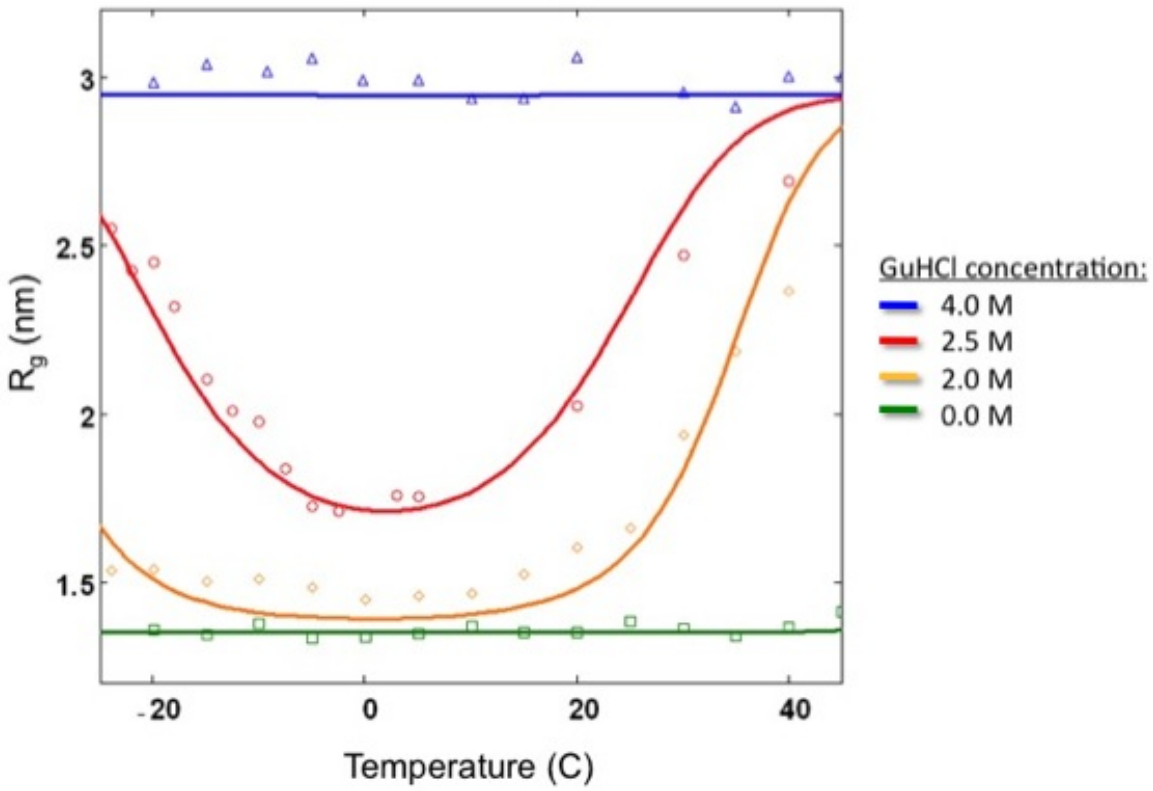


Figure 5.13: Both theoretical and measured Radius of Gyration verse Temperature in Celsius for Cytochrome-C in 0 molar, 2 molar, 2.5 molar and 4 molar concentrations of Guanidine HCl.

Figure 5.13 successfully demonstrates the correlation between the newly developed theoretical prediction of denaturation and the experimental reality of cold denaturation measured in the laboratory. The addition of a electrostatic component to the thermodynamic protein folding theory generates a model that is remarkably close to experimentally measured protein folding data. Further, the concentration of data points in the region of cold denaturation is the first of it's kind to be observed using small angle x-ray scattering. This experiment verifies cold denaturation using a global probe, allowing for state change to be observed.

Figure 5.13 tells us many things about the experimental setup as well as the theoretically developed cold denaturation prediction. Because the only error measurable was due to linear fitting of the data, data deviations from the theoretical predictions must be caused by something else. The most likely cause of miscorrelation of theoretical and experimental alignment is due to sample preparation and associated parameters. Specifically, the incorporation of a low pH to induce protein unfolding may be the cause of deviations in the colder range of fig. (5.13). As noted before, it was found that at cold temperatures, when the samples were chilled at -20 Celsius and then exposed for data collection at temperatures above -20 Celsius, analysis revealed all unfolded structures in populations with expected folded conformations. Later research showed that at low pH, pH less than 4, Cytochrome-C assumes the molten globule state (Kumar 2006). The molten globule state is a native-like structure with rigidity in its conformation due to instability caused by the high acidity or low pH (Kumar 2006). Once in the molten globule state, a protein will not leave its conformation. Because of this, samples exposed to temperatures that denature the protein and then examined at temperatures that do not denature the protein led to measurement of the molten globule state. It is not guaranteed that data points used in figures

(5.9) and (5.13) were taken with samples stored at room temperature to avoid this state because the problem was discovered later in the experimental run.

Further inconsistencies in data in comparison to the theoretical model may come from the data preparation technique itself. Several individuals prepared the samples of the same molar denaturant and protein. The volume of a prepared sample was 110 µl. Working with small volumes and preparing individual samples by various individuals may be the cause of deviations from the theoretical model. In fig. (5.13) the protein sample in 4 molar denaturant is obviously unfolded in both the theoretical and experimental case. Even with deviations in radius of gyration values in the experimental data, the protein can be easily determined to be in the denatured conformation. The protein sample in 0 M concentration of Guanidine HCl also demonstrates that with slight deviations from the theoretical model, the protein is obviously in the native state. To determine the consistency of sample preparation, the protein must be analyzed using laboratory techniques that would destroy the sample before it could even be tested. It is possible to lower error though.

To improve experimental radius of gyration calculations, several basic laboratory techniques may be incorporated to reduce error in measurements. Increasing the pH of the solution to at least 5 would remove the possibility of the protein falling into the molten globule state at cold temperatures. This would directly improve the 2 molar and 2.5 molar radius of gyration calculations because at cold temperatures, Cytochrome-C begins to unfold in these solutions as figures (5.9) and (5.13) show. Further, sample preparation improved by moving from the individual sample preparation technique to a multiple sample preparation technique. Samples would be prepared in large quantities and then divided into the 110 µl volume needed to take data. This would allowing quality control testing by setting aside one of the samples created for further testing to determine protein and denaturant concentrations. This would also eliminate the error incurred by attempting to repetitively measure out volumes as small as 5 µl accurately. If implemented, these changes would be most obvious in the 0 and 4 molar denaturant samples and lead to higher confidence in the 2 and 2.5 molar readings.

Theoretical calculations performed under the assumption of a population of identical particles may be the cause of error in both the theoretical model and the theoretical derivation leading to the development of fitting in Igor Pro. Incorporation of a mixed population into experimental data analysis may improve measurement quality. Guinier analysis calculates an average radius of gyration. Guinier Analysis has other models for mixed populations of both size and shape. Expanding the derivation of the radius of gyration to consider multiple radii may allow for a mixed population to be incorporated into measurement. Reduced scattering data for 2 molar and 2.5 molar denaturant samples demonstrated variation in protein sample. After data was reduced and plotted on a log-log scale, the protein samples associated with these molarities of denaturant lacked consistent scattering curves. Before the background could be subtracted and data Guinier analyzed, these curve derivations made data reduction difficult. Incorporating a multiple mean radii of gyration relation would make sample analysis more clear and potentially eliminate errors in analysis. Finding another way to tackle these curve derivations would drastically improve derivations from the theoretical model for 2 and 2.5 molarity curves.

The two state model was used in both theoretical derivation and incorporation in theoretical modeling and program development in Igor Pro. Because it is known that Cytochrome-C expresses two intermediate states as well as the final native or denatured states, an incorporation of multiple radius of gyration per sample may resolve any deviations in modeling as well as sample interpretation for the 2 and 2.5 molar data.

Improvements in the theoretical model could also be made to better describe the denaturation of Cytochrome-C. In this thesis, the two known intermediate states of Cytochrome-C were ignored and the protein was treated as if it had only a native and denatured state. Incorporation of the intermediates into the population model may allow for more accurate predictions. Future work may incorporating these intermediates may make the theoretical model better describe how the protein unfolds. Yet, Fig. 5.13 suggests that this incorporation is not necessary. Even with the error discussed in the experimental data, the two state model appears to fit Cytochrome-C very well. Improving sample preparation first and recomparing the theoretical to the experimental work would reveal if the two state model accurately describes Cytochrome-C's denaturation. For this thesis, the two state model is more than sufficient in accurately correlating physical parameters to the associated radius of gyration of Cytochrome-C.

Further measurements could be made of the physical variables involved in modeling to improve theoretical modeling of Cytochrome-C. The dielectric constant and pH of the solution as well as pk_i values of Cytochrome-C are functions of temperature. Measurement of the dielectric constant of the solution as a function of temperature has proven to be difficult and a curve fit to the dielectric constant of water as a function of temperature was used to calculate the theoretical model. Appendix 5 contains the dielectric fit for the sample solution using the assumption that the solution used has an associated dielectric constant that behaves exactly like water. A linear fit was used to approximate the dielectric constant as function of temperature. Modeling of the pH as a function of temperature can also be found in Appendix 5. The temperature dependent pH relation was established by taking pH measurements over a temperature scale of the actual buffered solution and then performing a linear fit to be used in theoretical modeling. Calculation of pk_i values for Cytochrome-C as a function of temperature requires destruction and testing of individual components of the protein and measurement of the associated pH at varying temperatures. As of now, the general temperature dependence for a given pk_i value is unknown and the laboratory lacks the facilities needed to perform the protein decomposition. The pH of the solution was measured and curve fit and then incorporated into the calculation of the theoretical model. Retesting of the temperature dependence of the pH of the solution would reassure confidence in the curve fit used in the theoretical model. It is clear that many improvements could be made to both the theoretical model and experimental data to more accurately depict the denaturation of Cytochrome-C.

Chapter 6

Conclusion

Section 6.1: Conclusion

The purpose of this thesis was to verify the cold denaturation of Cytochrome-C experimentally. It was shown that the thermodynamic conditions of a system containing Cytochrome-C do not induce denaturation. Taking into account the chemical or structurally dependent thermodynamics of the particular protein, a neutral folding free energy was derived. Further, the neutral folding free energy verify that if the chemical composition of the molecule is considered in cooperation with the thermodynamics of the system, Cytochrome-C will denatures both at a cold temperature and a hot temperature when the folding free energy of the system is zero. The neutral folding free energy proves that cold denaturation of the protein is possible but its associated temperatures are too extreme to be achieved with our apparatus. An electrostatic folding free energy was then developed to incorporate the impact of the electrostatic conditions of the protein and its surroundings. To do this, Cytochrome-C was modeled as a sphere with a surface charge and its electric field and potential were found. Using the potential energy relation to force, the energy due to a sphere with surface charge was defined. To relate both this energy and the energy from thermal effects, the Bjurrem length was incorporated. This allowed for a summation of energies into a final relation. Knowing the folding free energy of the system could now allow for prediction of a cold denaturation point.

To verify cold denaturation experimentally for Cytochrome-C, a method for measuring change in the protein needed to be derived. Small angle x-ray scattering in collaboration with Guinier analysis provides the means to measuring the size or radius of gyration of Cytochrome-C. To develop small x-ray scattering, the scattering off of an electron was considered first. This scattering theory was expanded to two electrons than multiple electrons to develop a form function. At the limit of small angles, several small angle assumptions were made possible and this form function was reduced to a Fourier transform of the associated form factor. Performing the Fourier transform produced a form function that could now be squared to give the scattering intensity off a sample. Following the theory of André Guinier, the radius of gyration was solved for from the intensity by defining the form function as an intrinsic function of the size of the protein and the angle of scattering. The conformation that the radius of gyration could be found by measuring the intensity scattered off of a protein made experimental small angle x-ray scattering of Cytochrome-C possible.

Once experimental time was allocated to expose Cytochrome-C to intense x-rays, experimental work commenced. This involved creating a solution that would be liquid below the freezing point of water to a certain degree as well as choosing denaturation conditions that would guarantee protein denaturation in the laboratory setting. Following the work of others, a low pH was chosen that would later prove to be problematic at cold temperatures. Guanidine HCl was utilized as the variable denaturant and a solution was conceived of ethylene glycol and water that would allow for examination of Cytochrome-C at temperatures below 0 Celsius. Modifications to an existing experimental setup had to be made to accommodate low temperatures.

Once data collection began, the data reduction and analysis program Igor Pro in cooperation with Macros created specifically for Guinier analysis by Liang Guo were used to determine the radius of gyration of all data samples. Acceptable error limits to the fitting parameters of Guinier analysis were imposed dependent on known values for the radius of gyration of Cytochrome-C in its various stages of denaturation. These error limits have made it impossible to contribute any sizable error to Guinier analysis.

The data now needed to be compared to the theoretical model. To do this, a two-state radius of gyration population model was used that incorporated the folding free energy of the protein and its surroundings. This model has made it possible to relate a theoretical model for protein denaturation to experimentally measured protein denaturation data. Upon comparison, it was found that the theoretical model does fit the experimental data. Adjustments in both experimental technique and theoretical modeling could be made to eliminate the deviations or errors between both.

Finally, it can be concluded that it is possible to reliably cold denature Cytochrome-C at temperatures below freezing. Further, a theoretical model has verified the experimental results found. Implementing the adjustments in technique could potentially render a highly accurate model verified very closely by actual experimental data. This work could easily be modified to incorporate other proteins given the right conditional information be provided. Future work may take these static studies of protein unfolding and use them to understand how the physical characteristics of the host cell lead to equilibrium protein folding and unfolding naturally.

While Cytochrome-C was a basic model for an intense study of cold denaturation, it makes way for study of larger and more important protein study. Further, understanding the physical parameters that maintain protein stability and instigate protein instability and unfolding lead to answering the rest of the protein folding problem.

Section 6.2: Future Work

Future work will commence on July 19th of 2010 to further research into cold denaturation of Cytochrome-C. After error analysis was considered, several changes to procedural techniques in sample preparation will take place and static data will be retaken on Cytochrome-C at a higher pH to avoid the molten globule state. Once the theoretical curve for each of the four molar denaturant concentrations has been reverified, more exciting techniques in data sampling will be implemented.

Using the perfected static cold denaturation curve for Cytochrome-C, the point of most denatured state will be identified. This temperature will be the starting point for time resolved studies of protein refolding (Gruebele 1998). To refold the protein on a time scale, an Neodymium-doped Yttrium Aluminium garnet, or Nd:YAG laser will be used to quickly raise the temperature of the sample. To do this, the Nd:YAG laser, which normally outputs light at 1064 nm will be raman shifted using a pipe filled with methane to 1.5 microns (Spuler 2007). At 1.5 micron wavelength, the OH bond of the water molecule becomes excited and energy absorption occurs. Thermal transfer in solution should induce protein refolding. Results using other observational techniques have already shown that this method works to refold a protein (Gruebele 1998). A high speed shutter and fast x-ray detector along with the Advanced Photon

Source's x-ray pulse pattern will allow for short exposure scattering patterns to be taken as the protein refolds from laser induced heating.

After sampling has occurred, data reduction again through Igor Pro should provide the data files needed to reconstruct a time dependent radius of gyration change like that introduced and discussed in this thesis.

If this work is successful, this procedure could be applied to studying the refolding of other proteins. Further, the work could verify the computational modeling discussed in Chapter 1 and further answer aspects of the protein folding problem. This future work will provide insight into the time scale for each aspect of the computational model leading to better understanding of the order in which a protein assembles itself during or after assembly by the ribosome.

Works Cited

Akiyama, Shuji, and Satoshi Takahashi. "Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering." PNAS. 99.3 (2002): 1329-1334. Print.

Becktel, Wayne, and John Schellman. "Protein Stability Curves." Biopolymers. 26. (1987): 1861. Print.

Bilsel, Osman, and C Robert Matthews. "Molecular dimensions and their distributions in early folding intermediates." Current Opinion in Structural Biology. 16. (2006): 86-93. Print.

"BioCAT Links Page." BioCAT. BioCAT, 22 Mar 2009. Web. 11 Jun 2010. <http://www.bio.aps.anl.gov/links.html>.

"Cytochrome-C." Cytochrome-c - Arabian Camel. Web. 5 Jun 2010. <http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=protein&dopt=GenPept&list_uids=65443>

Dill, Ken, and Sabrina Bromberg. Molecular Driving Forces. New York: Garland Science, 2003. 142-144. Print.

Dobson, C.M., Sali, A., and Karplus, M. (1998) Protein folding: a perspective from theory and experiment. Angew.Chem. Int. Ed. Eng., 37, 868.

Fischetti, R, S Stepanov, G Rosenbaum, R Barrea, and E Black. "The BIOCAT undulator beamline 18ID: a facility for biological non-crystalline diffraction and x-ray absorption spectroscopy at the Advanced Photon Source." Journal of Synchrotron Radiation 11.(2004): 399-405. Web. 10 May 2010. <http://www.bio.aps.anl.gov/scihi/2004-JSR.pdf>.

Ghosh, Kingshuk, and Ken A. Dill. "Computing protein stabilities from their chain lengths." PNAS 106.26 (2009): 10649-10654. Web. 18 Jan 2010. <www.pnas.org/cgi/doi/10.1073/pnas.0903995106>.

Glatter, O, and O Kratky. Small Angle X-ray Scattering. New York: Academic Press, 1982. 3-390. Print.

Gruebele, Martin, and Jobiah Sabelko. "Laser Temperature Jump Induced Protein Refolding." Acc. Chem. Res. 31. (1998): 699-707. Print.

Hsu, IJ, and YJ Shiu. "A solution study on the local and global structure changes of cytochrome c: an unfolding process induced by urea." J Phys Chem A 111.38 (2007): 9286-90. Web. July 9, 2009.

Kumar, R. "The Alkali Molten Globule State of Horse Ferricytochrome c: Observation of cold denaturation." J Mol Biol. 3.363 (2006): 483-95. Web. 23 Dec 2009. <http://www.ncbi.nlm.nih.gov/pubmed/17027030>.

Nelson, David L., and Michael M. Cox. Principles of Biochemistry. 4th ed. New York: W.H. Freeman and Company, 2005. 146-154. Print.

Pollack, Lois, and Mark W. Tate. "Compactness of a protein measured by sub-millisecond small-angle x-ray scattering." Proc. Natl. Acad. Sci. Usa. 96. (1999): 10115-10117. Print.

Pain, Roger H. Mechanisms of Protein Folding. New York: Oxford University Press, USA, 1994. 1-47,137-151. Print.

Press, HighWire, and Roger Pain. Mechanisms of protein folding. 2nd ed. Avon, United Kingdom: Oxford University Press, USA, 2000. 1-144. Print.

Privalov, P.L., and N.N. Khechinashivili. "A thermodynamic Approach to the Problem of Stabilization of Globular Protein Structure: A Calorimetric Study." Journal of Molecular Biology. 86. (1974): 665-684. Print.

Privalov PL ((1990) Biochemsitry and Molecular Biology 25.4 290

RCSB, PDB. "Structure Summary for Horse Cytochrome-C." RCSB Protein Data Bank. Rutgers,UCSD, Web. <http://www.pdb.org/pdb/explore/explore.do?structureId=1HRC>.

Robertson, Andrew D, and Kenneth P Murphy. "Protein Structure and the Energetics of Protein Stability." American Chemical Society. 96.383-4 (1997): Print.

Rodger A. and B. Norden (1997). Circular Dichroism and Linear Dichroism. Oxford University Press.

Roe, Ryong-Joon. Methods of X-ray and Neutron Scattering in Polymer Science. 1. New York City: Oxford University Press, USA, 2000. 150-165,302-306. Print.

Silvius, J.R. and Nabi, I.R. Fluorescence-quenching and resonance energy transfer studies of lipid microdomains in model and biological membranes. (Review) Molec. Membr. Bio. 2006, 23, 5-16.

Spuler, Scott M., and Shane D. Mayor. "Raman shifter optimized for lidar at 1.5 micron wavelength." Applied Optics. 46.15 (2007): 2990-2995. Print.
"Table 1. Dielectric constant of water ." Integrated publishing, tpub.com. tpub.con, n.d. Web. 21 Jan 2010. <http://www.tpub.com/content/ArmyCRREL/SR98_02/SR98_020009.htm>.

Wisz MS, Hellinga HW(2003) An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants. PROTEINS 51:360-377

Wu, Ying, and Elena Kondrashkina. "Microsecond acquisition of hetergeneous structure in the folding of a TIM barrel protein." PNAS. 105.36 (2008): 13367-13372. Print.

Appendices

Appendix Materials for Chapter 2:

Charged Group for Cytochrome-C	pK_{exp}	Reference
His18	2.40	Wisz 2003
His26	2.90	Wisz 2003
His33	6.40	Wisz 2003
Tyr67	11.00	Wisz 2003
Lys79	9.00	Wisz 2003

Table A.1: pK_i values needed to calculate localized protein charge, Q, for Cytochrome-C used to construct Fig. 5.10 using the complete folding free energy equation derived in chapter 2 and the population model derived in chapter 5.

Code corresponding to Images in chapter 2.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Graphing thermodynamic theory from Sara
% file name: theoretical_imaging.m
% created 6-14-2010
% file reproduces images relating H(T),S(T), Cp and more!
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

clear all

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Cp = 5000;%j/k/mol
T = 250:500; %kelvins
Th = 373; % kelvins using ITP
Ts = 385; % kelvins using ITP
y=-.8e6:1e4:.8e6;
HTh= 500000;% JOULES/MOLE, APPROXIMATELY WHAT WATER IS

HT = Cp.*(T-Th); % where it is assumed H(Th) is zero

figure(1); clf;
plot(T,HT,T,0,273,y);
xlabel('Temperature in Kelvins')
ylabel('Heat of Enthalpy in Joules');
title('The Heat of Enthalpy v. Temperature');
```

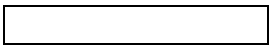
```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
ys=-4e3:10:2e3;

ST=Cp.*log(T./Ts);%entropy of the system

figure(2); clf;
plot(T,ST,T,0,273,ys);
xlabel('Temperature in Kelvins')
ylabel('Heat of Entropy in Joules per Kelvin');
title('Entropy v. Temperature');


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% To find the gibbs free energy, sum the entropy and the enthalpy

G = HT-T.*ST

yg = -2.5e5:1e3:1e5;

figure(3);clf;
plot(T,G,T,0,273,yg)
xlabel('Temperature in Kelvins')
ylabel('Gibbs free energy in Joules');
title(' Gibbs free energy v. Temperature');


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% now graph the gibbs free energy in parts

figure(4);clf;
plot(T,T.*ST,T,HT,T,0,273,y);
xlabel('Temperature in Kelvins')
ylabel('Energy in Joules');
title('Components of Gibbs free energy v. Temperature');
legend('Temperature scaled Entropy','Enthalpy');legend('boxoff');

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Now graph G,H,and T*S

figure(5);clf;
plot(T,T.*ST,T,HT,T,G,T,0,273,y);
xlabel('Temperature in Kelvins')
```

```
ylabel('Energy in Joules');
title('Enthalpy, Entropy Scaled and Gibbs free energy v. Temperature');
legend('Temperature scaled Entropy','Enthalpy','Gibbs free Energy');legend('boxoff');

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Appendix Materials for Chapter 3:

Derivation of Fourier Transform of a sphere:

Given the form factors in equations (3.18) and (3.19), the integral for the form function of the sphere can be performed. To do this, start by making the assumption that small angle x-ray scattering occurs in the Fresnel or far field. This assumption allows the dot product describing the phase shift of the scattered x-rays to reduce to a scalar,



(A3.1)

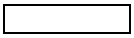
. The integral form of the form function may now be expressed in terms of the form factor

(A3.2)

or, incorporating the form function inside the sphere,

(A3.3)

(Roe 2000). The form factor outside of the sphere disappears and will be ignored in this calculation. A variable substitution can be used to simplify the integrals by letting



(A3.4)

. This substitution reduces the form function yet again

(A3.5)

. Taking the first two integrals generates a form function in terms of a sum of exponents

(A3.6)

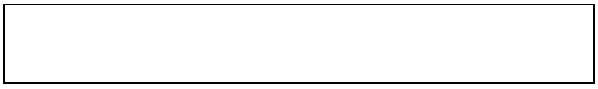
. The inside of this integral looks exactly like the exponential form of a sine function. A trigonometric substitution into the integral will again reduce the form function into something calculable

(A3.7)

. Integration by parts gives the final form of the form function in terms of Q and R and completion of the above integration,

(A3.8)

. The integral generates a form function in terms of the radius and the scattering vector,



(A3.9)

. Applying the limits of the integral and separating out a volume term will generate the final form function for a sphere of scattering surfaces,

(A3.10)

. Squaring the form function will generate a relationship for the intensity of the scattered x-ray beam off the surface of a sphere of radius R. Scaling by the appropriate density function will generate a relationship for the intensity in terms of the scattering vector Q.

Appendix Materials for Chapter 4:

pH meter Calibration

1) Monitor pH for each protein solution (from http://www.chemistry.nmsu.edu/studntres/chem435/Lab12/accumet.pdf)
    a. With meter turned on, make sure electrode in mV measuring mode.
    b. Place electrode in pH 4 buffer and read
    c. Place electrode in pH 7 buffer and read
    d. If required, place in pH 10 solution. Take reading.
    e. Finally, rinse electrode or wipe with kimwipe/cleanwipe and place in solution to be measured.
    f. Take reading, record in book.
        i. Should be pH 3.5 approximately, if not, and decrease pH to desired measurement by adding HCl drop by drop and taking measurements intermittently

Sample Loading

1) Load appropriate sample (for both protein and background samples)
    a) **Make sure pump is not running! AND plunger is at top of cycle.**
    b) Open background cuvette and place inside sample chamber. **The tube should just touch the bottom of the cuvette.**
    c) On Microlab pump controller select existing program
    d) From menu of existing programs, select GL-Protein
    e) Confirm twice (it asks if you are really sure you want to run the program which of course you do)
    f) Select run
    g) Ensure monitor camera is focused on pump controller screen
    h) Search and close hutch
    i) Open shutter
    j) Check file name corresponds to sample loaded
    k) Trigger program

Imaging and Storing Data

1) Correctly name files on corresponding computers, syncing and triggering data collection.
    a) On both SAXS collection and data processing computer, ensure both directories show F:\SAXS_logs\2009run3\Landahl121609_____
    b) The blank part will be the new log file, it should follow this template: m025c3:

1. The number next to the m is the molarity of denaturant (so this example is 0 molarity denaturant).
2. The 25 is the temperature in Celsius, for negative temperatures, use _, so -5 reads _5.
3. The final number (to the right of the c) is the trial number starting at 1.
4. The template therefore stands for the third trial for 0 molar GuHCl at 25^{o}C

2) Syncing file names and preparing to collect data
   a) On processing computer, enter file name and hit enter/play?
   b) after 5 seconds of elapsed time, a program button will indicate that the computer is ready to take data
   c) ready computer for data collection by clicking on play
3) Starting data collection
   a) After processing computer is prepared, check both computers to ensure file names are IDENTICAL.
   b) Trigger pump and start program collection program on SAXS collecting computer at relatively the same time.

Capillary Cleaning Procedure

- After testing the buffer solution only:
  - Allow two full pumps of 45/55 mix to go through the capillary (the pump must go up and down twice)
  - Allow all solution to evacuate before setting up next sample
- After testing protein solution
  - Two pumps 45/55 mix
  - Two pumps isopropyl alcohol
  - Two pumps 45/55 mix
  - Two pumps ___% bleach solution
  - Two pumps 45/55 mix
  - Allow all solution to evacuate before setting up next sample

When ceasing the experiment for an extended amount of time, allow one pump of bleach solution to flow into capillary and then turn off pump before it is evacuated.

Coordinating File Locations for Data Analysis

1) Moving files to appropriate directory
   a. Determine correct directory for data reduction
      i. Files should originate in _____
      ii. And be manually moved to _____

b. Before data reduction, make sure mask is in same directory as data collected and igor pro

Data Reduction Instructions

A Comprehensive Outline to Saxs Data Reduction

Purpose: Using Igor Pro in conjunction with macros provided by Liang Guo, to reduce data taken on 16-21 December 2009 and discern the Radius of Gyration of Cytochrome C in different concentrations of denaturant and at different temperatures.

Goal: Phase 1 will reproduce and edit existing Radius of Gyration calculations. Finally, use data to determine where to repeat data acquisition.

(assuming macros have been replaced and edited for mac)

I. Reducing Data in Igor Pro

Open Igor Pro
BioCAT (on tool bar)
SAXS data reduction
Choose appropriate mask
NOTE: s121109.msk for data used in this experiment

To add or change possible mask options:

Data (On tool bar)
Load waves
Load Igor Text
In popped up window:
Enable: Igor Text File
Load mask (now choose .msk file from where stored)

I Continued reduction(1)

1) After choosing mask
2) Continue (do not change anything in this window!)
3) Check beam location (location stored in **notebook**)

How to access notebook

Windows (in tool bar)
Other windows
Notebook0

NOTE: all beam locations and associated masks are stored in the text file: **Params.txt**

I Continued reduction Again(2)

After verifying correct beam location, continue

1) Select the log file corresponding to desired data to be analyzed
2) Open
   Select first .tif file in data set
   Open

II Viewing reduction

Windows (on tool bar)
Command window (window will pop up and display reduction)

AFTER data is reduced, analysis can begin.

III Plotting raw data

1) Plot fits
   Loglog plot
   Loglogplot
2) Make sure number of data to plot is 30 or for some data sets, 45 or 60.
3) Set 'select the Q data:' to first file reduced or .tif file used to process reduction
   Continue

Note: If macro error appears, close out, don't fear
Loglog plot will appear.

IV Selectively removing Bad Data

NOTE: goal is to process two solid, separate readings (buffer and solution) and to get this by removing outlying data if needed.

1) To do this expand around fragmented areas by
2) Right click, click inside box, expand
3) Repeat until optimal viewing is achieved
4) Use '□A' to minimize back to original screen
5) To select outlying data for removal, Place mouse over wily line, right click, choose to remove from pop up menu
6) Repeat until both buffer and solution look coherent

NOTE: do not remove more than 67% of data from each

IF PLOT LOOKS MODERATELY ACCEPTIBLE: **CONTINUE**
if not, document and start next file.

V Subtracting out Background

a. save data lines separately:

1) Box lower, buffer, data plot line
   '□p'
2) rename file by adding '_buf' to end
   continue

3) Box in upper, solution, data plot line
‘□p’
keep name same- continue

b. subtract buffer from solution
1) ‘□q’
2) in ‘enter background data wave:’ select the buffer data just created
3) in ‘enter data wave:’ select solution data just created
continue

VI. Plotting subtracted data
This graph will reveal if data taken was good, bad, and folded or unfolded

1) Plot fits
Loglogplot
Loglogplot

2) Change Number of data sets to plot:’ from 30 to 1
3) Change ‘Select the Q data’ to ‘q(filename/handle)_avg_bsub
Continue
NOTE: Image will appear, from this, denaturation should be easily discerned.
4) Save experiment (this saves graph)

VII. Making a Guinier Plot
Plot fits
Special plots
Make Guinier plot
Change ‘enter the data to plot’ to file used to graph subtracted data from above
Continue

VIII. Measuring R_g using our Guinier Plot
a. Making the Guinier Plot
In newly produced Guinier graph, expand straight –ish region to the far left.
Double click line
Change ‘mode’ to ‘markers’
Do it!!

b. Performing a Guinier Fit
1) Using the tool bar that pops up with plot, click inside circle
2) Drag circle over line to furthest left end of straightest aspect of markers (may use left and right arrow on key board to move Circular cursor left or right along line)
3) Click square below circle and drag to line. Placeto the right of circle, on line. These steps may be repeated to test Rg
4) plot fits
special plots

perform Guinier fit

5) 'enter the data to fit:' should be set to the subtracted file name with the prefix 'rgdp_'
continue

Things to consider,

A) $1.3 > Q_{max}*R_g > 1$, if it is out of this region, repeat steps 1-5 until within these limits.
B) note how Rg changes with refitting. In some cases, Rg is nicer when Q*R is closer to 1 or closer to 1.3.
C) watch out for aggregation

Finally!!! Record Rg in word table, save experiment, close first graph and save experiment please!

Appendix Materials for Chapter 5:

Temperature in degrees C	Radius of Gyration in Angstroms	Fit error in □ Angstroms	□ in Angstroms^2
-20	13.59	.05	1.27
-15	13.45	.08	1.14
-10	13.76	.08	1.16
-5	13.37	.03	1.33
-5	13.56	.56	1.28
0	13.38	.04	1.29
0	13.94	.78	1.30
5	13.47	.05	1.22
10	13.69	.04	1.28
10	15.82	.00	1.21
15	13.49	.06	1.22
20	13.50	.04	1.30
20	13.62	.10	1.01
25	13.85	.14	1.25
30	13.62	.07	1.23
35	13.38	.04	1.13
40	13.66	.08	1.08
45	14.10	.06	1.32
50	14.06	.06	1.19

Table A5.1:Cytochrome-C in 0 Molar Guanidine HCl solutions radius of gyration measurements including error and "QR" values for data taken during experimentation and analyzed using Igor Pro.

Temperature in degrees C	Radius of Gyration in Angstroms	Fit error in □ Angstroms	$Q_{max} \cdot R_g$ in Angstroms^2
-24.5	15.34	0.11	1.17
-20	15.38	0.08	1.18
-15	16.69	0.17	1.10
-10	15.09	0.09	1.12
-5	14.85	0.09	1.23
0	14.49	0.06	1.31
5	14.57	0.11	1.11
10	16.69	0.39	1.29
15	15.47	0.09	1.20
20	15.44	0.09	1.16
25	16.36	0.09	1.19
30	19.36	0.10	1.30
35	21.83	0.21	1.24
40	23.63	0.26	1.31
49	24.01	0.14	1.45
50	25.01	0.19	1.24

Table A5.2:Cytochrome-C in 2 Molar Guanidine HCl solutions radius of gyration measurements including error and "QR" values for data taken during experimentation and analyzed using Igor Pro.

Temperature in degrees C	Radius of Gyration in Angstroms	Fit error in ± Angstroms	$Q_{max} \cdot R_g$ in Angstroms^2
-24.5	25.50	0.01	1.06
-22.5	24.24	0.36	1.20
-20	24.50	0.33	1.10
-18	23.17	0.39	1.04
-15	21.02	0.18	1.23
-12.5	20.08	0.20	1.22
-10	19.75	0.20	1.24
-7.5	18.35	0.18	1.27
-5	17.25	0.10	1.12
-2.5	17.11	0.21	1.22
3	17.58	0.23	1.14
5	17.54	0.21	1.25
20	20.23	0.17	1.27
30	24.70	0.34	1.19
40	26.91	0.28	1.21

Table A5.3:Cytochrome-C in 2.5 Molar Guanidine HCl solutions radius of gyration measurements including error and "QR" values for data taken during experimentation and analyzed using Igor Pro.

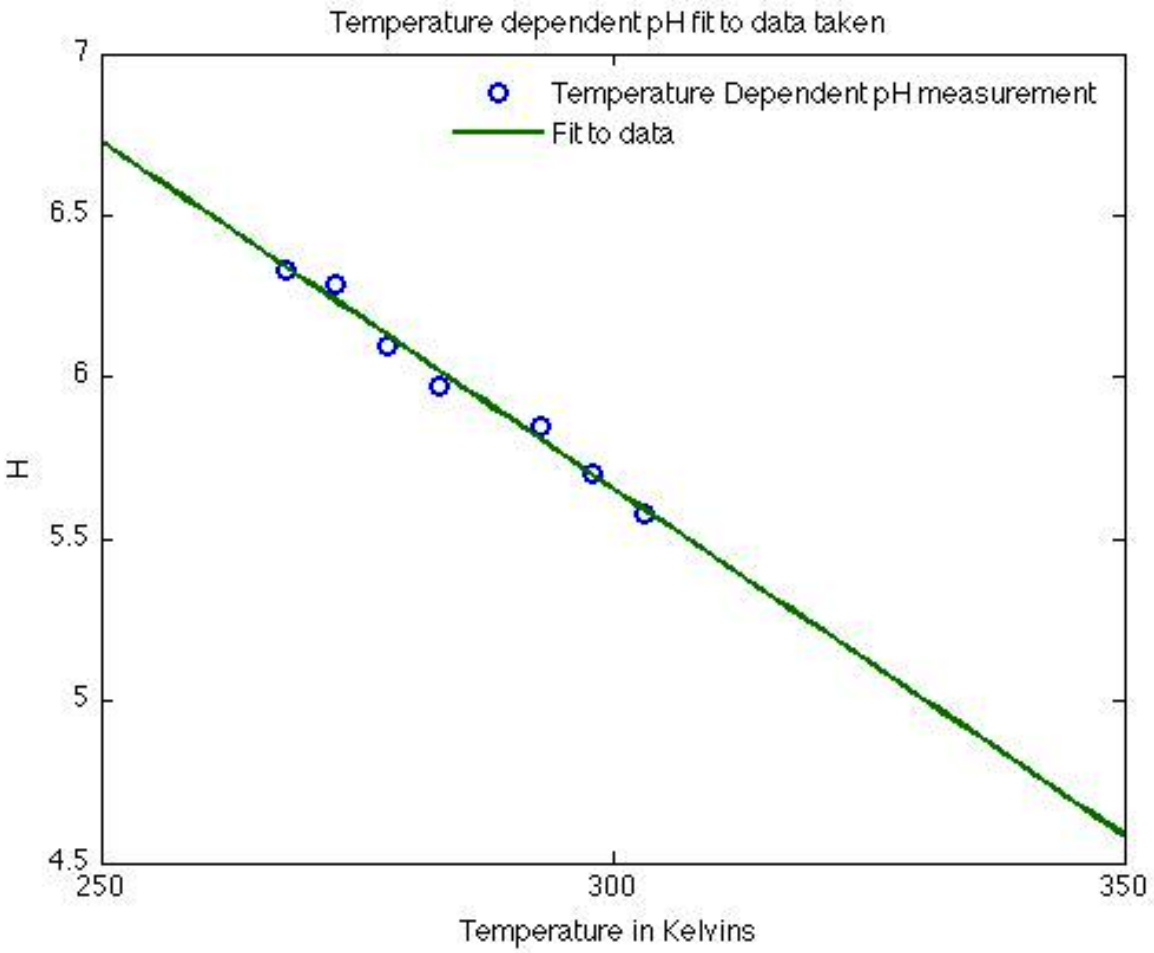
Temperature in degrees C	Radius of Gyration in Angstroms	Fit error in ± Angstroms	$Q_{max} \cdot R_g$ in Angstroms^2
-20	29.86	0.82	1.34
-15	30.39	0.87	1.28
-10	30.19	0.41	1.27
-5	30.57	0.93	1.24
0	29.91	0.78	1.30
5	29.94	0.69	1.30
10	29.40	0.69	1.32
15	29.36	0.54	1.28
20	30.61	0.72	1.29
30	29.55	0.77	1.28
35	29.12	0.62	1.31
40	30.03	0.49	1.31
45	30.03	0.45	1.31
50	30.66	0.72	1.29

Table A5.4:Cytochrome-C in 4 Molar Guanidine HCl solutions radius of gyration measurements including error and "QR" values for data taken during experimentation and analyzed using Igor Pro.

Code mapping data from tables A5.1-5.4

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%
% Plotting Rg data by molarity
% file name: Rg_plot.m by Margaret Elmer
% last editted 7-4-2010
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%
clear all;

t0 = [-20 -15 -10 -5 0 5 15 20 25 30 35 40 45 50];

rg0= [13.589 13.448 13.763 13.37 13.377 13.472 13.492 13.5
13.848 13.619 13.383 13.657 14.102 14.056];
```

```
t2 = [-24.5 -20 -19.4 -19 -15 -10 -5 -.2 0 5 9.5 10 15 20 25 30
35 40 49 50];

rg2 = [15.138 15.378 16.689 15.435 15.01 15.094 14.849 15.265
14.494 14.573 13.429 14.667 15.245 16.016 16.582 19.363 21.834
23.634 24.011 25.004];

t25 = [-24.5 -22.5 -20 -18 -15 -12.5 -10 -7.5 -5 -2.5 2 5 15 20
25 30 40];

rg25 = [25.5 24.236 24.542 23.165 21.022 20.08 19.752 18.348
17.253 17.105 17.576 17.535 19.848 20.23 24.88 24.699 26.912];

t4 = [-20 -15 -10 -5 0 5 10 15 20 30 35 40 45 50];

rg4 = [29.862 30.39 30.185 30.566 29.905 29.935 29.396 29.361
30.611 29.545 29.123 30.033 30.033 30.657];

figure(1); clf;
size =60;

set(gcf,'color','w');

 scatter(t0, rg0,size,'r')
 hold on
 scatter(t2, rg2,size,'g')
 scatter(t25,rg25,size,'b')
 scatter(t4, rg4,size,'m');
 hold off
```



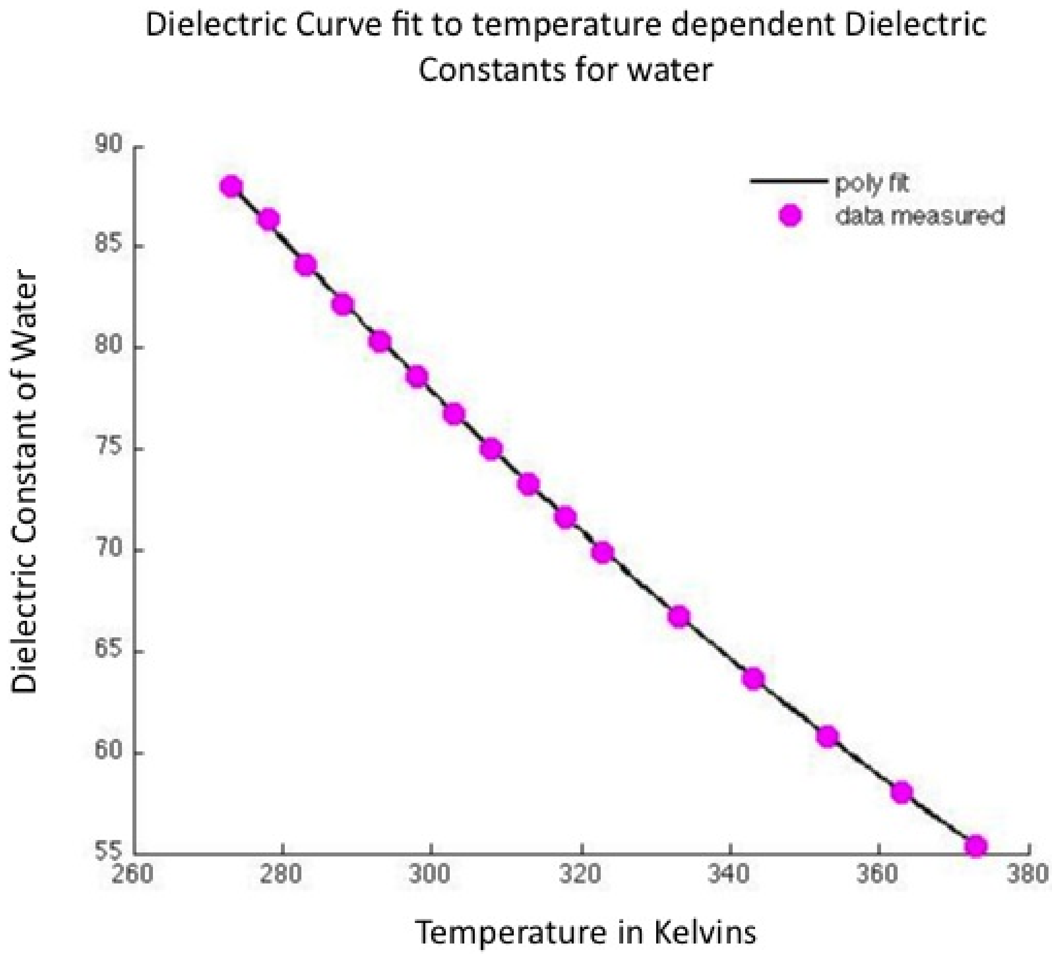
A5.5: Mathematical fit to pH dependent temperature measurements taken on samples of 45/55 buffered mix at varying temperatures.

Equation fitting pH measurements of sol45/55 buffered mix at varying temperatures

Code producing fitting parameter for pH(T) for 45/55 buffered mix

```
temp1 = [-5+273 273+0 273+5 273+10 273+20 273+25 273+30]
ph1 = [6.31 6.29 6.09 5.96 5.85 5.71 5.58]
temp2 = [-5+273 273+0 273+5 273+10 273+20 273+25 273+30]
ph2 = [6.35 6.29 6.11 5.98 5.85 5.69 5.58]

%temp = [-5+273 -5+273 273 273+0 273+5 273+5 273+10 273+10
273+20 273+20 273+25 273+25 273+30 273+30];
%ph = [6.31 6.35 6.29 6.29 6.09 6.11 5.96 5.98 5.85 5.85 5.71
5.69 5.58 5.58];
```

```
temp = (temp1+temp2)./2;
ph = (ph1+ph2)./2

x = 250:1:350;

p = polyfit(temp,ph,1);
b = p(2)
m = p(1)

y1 = m.*x+b;

pp = polyfit(temp,ph,2);
c = pp(3);
b = pp(2);
a = pp(1);

y2 = a.*x.*x+b.*x+c;

ppp = polyfit(temp,ph,3);
g = ppp(4);
f = ppp(3);
e = ppp(2);
d = ppp(1);

y3 = d.*(x.^3)+e.*(x.^2)+f.*x+g;

figure(1);clf;
plot(temp,ph,x,y1,'LineWidth',2)

ph = ph-2.2; % to get pH of 3.5
x = 250:1:350;

p = polyfit(temp,ph,1);
b = p(2)
m = p(1)

xlabel('Temperature in Kelvins');
ylabel('Hp')
set(gcf,'color','w');
title('Temperature dependent pH fit to data taken');
Legend('Temperature Dependent pH measurement','Fit to data');
Legend('boxoff');
y1 = m.*x+b;
```



A5.6: Dielectric fit to liturature values of the dielectric constant of water at varying temperatures to emulate the dielectric constant using throughout experimentation.

The associated fit equation for using the matlab code below is:

Code for dielectric constant fitting and figure

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% finding the dielectric constant of water at different temperatures
% file name: dielectric_calculator.m
% last editted 2-8-2010 by margaret elmer
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%%%%%%%%%
clear all;

temp = [273+0 273+5 273+10 273+15 273+20 273+25 273+30 273+35
273+40 273+45 273+50 273+60 273+70 273+80 273+90 273+100];
% temperature in 5 degree celcius increments
dielectric = [88 86.40 84.11 82.22 80.36 78.54 76.75 75 73.28
71.59...     % dielectric constant of water from wikipedia;
remember this is a dimensionless quantity
    69.94 66.74 63.68 60.78 57.98 55.33];

p = polyfit(temp,dielectric,1);
% polyfit a line and graph over the index
b = p(2);
m = p(1);

x = 273:1:373;
% create index to graph line over
y = m.*x+b;

p2 = polyfit(temp,dielectric,2)
% polyfit polynomial to dielectric over index
y2 = p2(1).*x.^2+p2(2).*x+p2(3);
size = 50;

figure(1); clf;
set(gcf,'color','w');
hold on
plot(x,y2,'color','k','LineWidth',2);
plot(temp,dielectric,'color','m','LineWidth',5);
hold off
legend('poly fit','data measured');legend('boxoff');
```